

C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
 F:658-680/Domain: transmembrane #status predicted <TMN>
 F:723-988/Domain: protein kinase homology <KIN>
 F:731-739/Region: protein kinase ATP-binding motif
 F:71-191,263,535,576,634/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:691/Binding site: phosphate (Thr) (covalent) #status predicted
 F:758/Active site: Lys #status predicted
 F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.4%; Score 5918; DB 1; Length 1260;
 Best Local Similarity 86.4%; Pred. No. 4,8e-234; Indels 14; Gaps 5;
 Matches 1092; Conservative 51; Mismatches 105;

QY 1 MELAAACRWGLLLALLPGCAASTQVCTGTDMLRLPASPETHDMLRLHYQCCVQVQNL 60
 DB 4 MELAAACRWGLLLALLPGIAGTQVCTGTDMLRLPASPETHDMLRLHYQCCVQVQNL 63
 QY 61 ELTYLPTNASLFLQDIEQVQYVLIHNOVQVPLQRLIRVGTQLFEDNVALAVLNG 120
 DB 64 ELTYLPANASLFLQDIEQVQYVLIHNOVQVPLQRLIRVGTQLFEDKVALAVLNR 123
 QY 121 DPLNNTTPTV-GASPGGLRELQRLSLTEILKGVLIORNPOLCYODTILWKDF---NNFT 176
 DB 124 DPQDVAASTPRTPEGLRELQRLSLTEILKGVLIORNPOLCYODMVLWKDFVFRKN-- 181
 QY 177 VSWLVRPKVSASHLENRSRACHPCSPMKGSRGWGESSEDCQSILTRTVACGGCARCKGP 236
 DB 182 ----QLAPVDID--TNRSRACPPCAPACKONHGWGESPEDCQILITGICTSGCARCKGR 234
 QY 237 LPTDCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPPEGRYT 296
 DB 235 LPTDCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPPEGRYT 294
 QY 297 FGASCVTAPNYLSTGVSCCTLVCLHNOEVTADGTQCEKSKPCARVCYGLGMEHL 356
 DB 295 FGASCVTTCPNYLSGVSCCTLVCPNNQEVTAEDGTQCEKSKPCARVCYGLGMEHL 354
 QY 357 REVRAVTSANTQEFAGCKKIFGSLAFPEFSDGDPASNTAPLOPQLOVFTLEITGYL 416
 DB 355 RGAARITSDNVOEPDGCKKIFGSLAFPEFSDGDPSSGIAPLRPQLOVFTLEITGYL 414
 QY 417 YISAWPDSLPLSVFQNLQVTRGRILHNGAVSLTQGLIGISWLGRLSRLGSLALIH 476
 DB 415 YISAWPDSLRLSVFQNLRIIRGRILHDGAYSLTQGLIGISWLGRLSRLGSLALIH 474
 QY 477 NTHLCFVHTVPMDQLFRNPHQALLHTANRPEDE-CVGRGLACHOLCARGHCWGPCTOCV 535
 DB 475 NAHLCFVHTVPMDQLFRNPHQALLHSGNRPEDELCVSSGLVCNSLCAHGCWGPCTOCV 534
 QY 536 NCSQFLRQCEVEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADOCVACAH 595
 DB 535 NCSHFLRQCEVEECRVWKGKLPREYVSDKRLCPCHPECPQNSSETCFGSADQCAACAH 594
 QY 596 YKDPFPCVAPCGVKPDLSTWYPTWKFDPDEGACQPCPINTHSCVDLDDKCPAEQAS 655
 DB 595 YKDSSSCVARCPGSKPDLSTWYPTWKFDPDEGICQPCPINTHSCVDLDERGCPAEQAS 654
 QY 656 PLTSIVSAVGLLWVVLGVVFGILIKRQOKIRKVTMRLLQETELVELPTSPGAMPNQ 715
 DB 655 PVTIATVEGLVFLILVGVVGLIKRRQOKIRKVTMRLLQETELVELPTSPGAMPNQ 714
 QY 716 AQMRILKETELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILD 775
 DB 715 AQMRILKETELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILD 774
 QY 776 EAYMAGVGSYVSRLLGICLTSTVQLVTQMPYGCLLDHYVRENRGRIGSODLLNWCQI 835
 DB 775 EAYMAGVGSYVSRLLGICLTSTVQLVTQMPYGCLLDHYVRENRGRIGSODLLNWCQI 834
 QY 836 AKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWM 895

835 AKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWM 894
 QY 896 ALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTI 955
 DB 895 ALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTI 954
 QY 956 DYVIMTWKMWIDSCRRPRFRELVSFSEFARMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1015
 DB 955 DYVIMTWKMWIDSCRRPRFRELVSFSEFARMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1014
 QY 1016 EDDDMGDLVDAEYLVPOGFECPDPAPACAGGMVHRRSSSTRSGGGLTLGLPSEEE 1075
 DB 1015 EDDDMGDLVDAEYLVPOGFECPDPPTFTGTSTAHRHRRSSSTRSGGGLTLGLPSEEG 1074
 QY 1076 APRSLAPSEAGSGSDVDFGDLGMAAKGLQSLTPHDPSPLOQRYSEDPTVPLPSETQGYVA 1135
 DB 1075 PPSPLAPSEAGSGSDVDFGDLGMAKGLQSLTPHDPSPLOQRYSEDPTVPLPSETQGYVA 1134
 QY 1136 PLTCSPPQYVNVQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVFAGG 1195
 DB 1135 PLACSPQYVNVQSEVQPPPLTPGGLPPVVRPAGATLERPKTTLSPGKNGVVKDVFAGG 1194
 QY 1196 AVENPEYLTPOGGAAPQHPHPPAPSPAFDNLYWDQDPPERCAPSTFKGTATENPEYL 1255
 DB 1195 AVENPEYLVPREGTASPPHPPAFSPAFDNLYWDQNSSEQGPSPFNFEETATENPEYL 1254
 QY 1256 GLDVPV 1261
 DB 1255 GLDVPV 1260

RESULT 3
 148161
 P-185 precursor - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C:Accession: I48161
 R:Nakamura, T.; Uehijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Iehika
 Gene 140, 251-255, 1994
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A:Reference number: I48161; MUID:94193007; PMID:7908275
 A:Accession: I48161
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1254 <RES>
 A:Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595
 C:Genetics:
 A:Gene: neu
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP
 F:718-983/Domain: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 86.4%; Score 5914.5; DB 2; Length 1254;
 Best Local Similarity 86.4%; Pred. No. 6.6e-234;
 Matches 1092; Conservative 59; Mismatches 100; Indels 13; Gaps 4;

QY 1 MELAAACRWGLLLALLPGCAASTQVCTGTDMLRLPASPETHDMLRLHYQCCVQVQNL 60
 DB 1 MELAAACRWGLLLALLSPGASGTQVCTGTDMLRLPASPETHDMLRLHYQCCVQVQNL 60
 QY 61 ELTYLPTNASLFLQDIEQVQYVLIHNOVQVPLQRLIRVGTQLFEDNVALAVLNG 120
 DB 61 ELTYLPANATLSFLQDIEQVQYVLIHNOVQVPLQRLIRVGTQLFEDKVALAVLNR 120
 QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGVLIORNPOLCYODTILWKDF---NNFTV 177
 DB 121 DPLDNVTATGRTPEGLRELQRLSLTEILKGVLIORNPOLCYODTILWKDFVFRKN--- 177
 QY 178 SFWLVRPKVSASHLENRSRACHPCSPMKGSRGWGESSEDCQSILTRTVACGGCARCKGP 237
 DB 178 ----QLAPVDID--TNRSRACPPCAPACKONHGWGESPEDCQILITGICTTAPRAVPAARL 231

238 PTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTF 297
232 PTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTF 291
238 GASCVTACPNYLYSTDVSGCTLVCPHNOEVTABDGTQRCCKSPCARVCVGLGMEHLR 357
232 GASCVTTCPNYLYSTEVSGCTLVCPHNNOEVTABDGTQRCCKSPCARVCVGLGMEHLR 351
358 EVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLQVFPETLEEITGYLY 417
352 GARAITSANIQEFACKKIFGSLAFLPESFDGNPSGTAPLTPEQLQVFPETLEEITGYLY 411
418 ISAMPDSDLVSFQNLQVIRGRIILHNGAYSITLQGLGISMGLRSLRSLGSLALIRHN 477
412 ISAMPDSDLVSFQNLVIRGRLVHDGAYSLALQGLGIRMLGLRSLRSLGSLVLIHRN 471
478 THLCFVHTVPWDLFRNPHQALLHTANRPEDCEVGEGLACHOLCARGHCWGPGPTQVCNC 537
472 THLCFVHTVPWDLFRNPHQALLHSGNPSEECGLKDFACYPLCAHGHGCGPPTQVCNC 531
538 SOFLRGQECVEBCRVLQGLPREYVVARHCLPCHPECPONGSVTCFPGPEADOCVACAHYK 597
532 SHFLRGQECVCKRWKGLPREYVNGKCLPCHPECPONSTETCTGSEADQCTACPHYK 591
598 DPFFCVARCPGKPDLSYMPKPFDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPL 657
592 DSPFFCVARCPGKPDLSYMPKPYDEEGMCQPCPINCTHSCVDLDERGCPAEQASPA 651
658 TSIVSAVVGILLVVLGVVFGILLIKRQOKIKRYTMRLLQTELVEPLTPSGAMPNQAK 717
652 TSIIATVVGILLVVLGVVVGILLIKRRQKIKRYTMRLLQTELVEPLTPSGAMPNQAK 711
718 MRILKETLRKVKVLGSGAFGVYAGIWPDGENVKIPVAIKVLRNTSPKANKILDEA 777
712 MRILKETLRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKILDEA 771
778 YVMAGVSPYVRLLGICLTSTVQLVTLQMPYGCILLDHWRENRGLSGODLLNWCQIAK 837
772 YVMAGLSPYVRLLGICLTSTVQLVTLQMPYGCILLDHWRENRGLSGODLLNWCQIAK 831
838 GMSYLEDLVRLHRLDAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMAL 897
832 GMSYLEDLVRLHRLDAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMAL 891
898 ESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDV 957
892 ESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDV 951
958 YMIWVKCWMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPSPDLSTFYRSLLED 1017
952 YMIWVKCWMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPSPDLSTFYRSLLED 1011
1018 DMGDLVDAEEYLVPOQGFPCDPARGAGCMVHRRSSSTRSGGDLTLGLEPSEEEAP 1077
1012 DMGDLVDAEEYLVPOQGFPPDPAPGAGSTAHRRSSSTRSGGDLTLGLEPSEEEAP 1071
1078 RSLAPSEGAGSDVFDGLGMAAGLQSLPDRPSLQRYSEDTVPPLPSETDGVVAPL 1137
1072 RSLAPSEGAGSDVFEGLGMCATKGQSI SPRDLSPLQRYSEDTLPPTETDGVVAPL 1131
1138 TCSPOPEYVNPDPVPPOPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGAV 1197
1132 ACSPOPEYVNPDPVPPOPLTPPEGLPLPVRPAGATLERPKTLSPGKNGVVKDVFTFGAV 1191
1198 ENPEYLTPOGGAPOPHPPAPSPAFDNLVYWDQPPPERGAPSTFKCTPTAENPEYLG 1257
1192 ENPEYLVFRGGASQFH-PPALCPAFDNLVYWDQPPSERGSPNTFEGTPTAENPEYLG 1250
1258 DVPV 1261
1251 DVPV 1254

RESULT 4
GCHUE
epidermal growth factor receptor precursor - human
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Homo sapiens (man)
C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999
C:Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143;
R:Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.
rg, P.H.
Nature 309, 418-425, 1984
A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression
A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <ULL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A>Note: The authors translated the codon AAG for residue 540 as Asn
R:Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R:Hailey, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identificatio
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Hailey, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termi
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A:Experimental source: carcinoma cell line A431-7
R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.;
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RN
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 'RCAWRRA', 150-187, 'KSVIOAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-
, 798-799, 'TD', 802-811, 'R', 813-942 <XUY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF re
R:Liu, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma,
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplificati
A:Reference number: A43615; MUID:84196372; PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
 C:Species: Gallus gallus (chicken)
 C>Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
 C:Accession: A27720; A00643
 R:Lax, J.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennart
 Mol. Cell. Biol. 8, 1970-1978, 1988
 A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
 A:Reference number: A27720; MUID:88261272; PMID:3260329
 A:Accession: A27720
 A:Molecule type: mRNA
 A:Residues: 1-1223 <LAX>
 A:Cross-references: GB:M20386
 R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
 Cell 41, 719-726, 1985
 A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
 A:Reference number: A00643; MUID:85228222; PMID:2988784
 A:Accession: A00643
 A:Molecule type: mRNA
 A:Residues: 585-1223 <NIL>
 A:Cross-references: GB:M10066
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
 specific protein kinase
 F:1-30/Domain: signal sequence #status predicted <SIG>
 F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
 F:31-654/Domain: extracellular #status predicted <EXT>
 F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
 F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
 F:655-677/Domain: transmembrane #status predicted <TM>
 F:678-1223/Domain: intracellular #status predicted <INT>
 F:719-984/Domain: protein kinase homology <KIN>
 F:727-735/Region: protein kinase ATP-binding motif
 F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #8
 F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:754/Active site: Lys #status predicted
 F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.4%; Score 3110.5; DB 1; Length 1223;
 Best Local Similarity 48.5%; Pred. No. 1.1e-119;
 Matches 632; Conservative 172; Mismatches 349; Indels 149; Gaps 25;

QY 8 RMGLLLALLPPCAA-----STQVCTGTDMLKRLPASPEHLDMLRHLYQGCQVQGNLE 61
 DB 13 RGAALVLLVLLGVALCSAVEEKVCQGNKLTQLGHVEDHFTSLQRMYNNECVVLSNLE 72
 QY 62 LTYLPTNASLSFLQDIQEVQGVLYIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNGD 121
 DB 73 ITYVHNRLDTFLKTIQEVAGVLYIALNMVDVPLENLQIRGNVLYDSFALAVLSNVH 132
 QY 122 PLNNTPTVTGSPGIGRELQRLSLTEILKGVLIQRLPOLCYQDILKDFNFTVFWL 181
 DB 133 -MNTQ-----GLRELPMKRLSEILNGGVKISNNPKLCNMVDVLMNDIIDTS----- 178
 QY 182 RPKVKSASHLENRSRACHPCSPMKGRCWGSSESDCSLRTVTCAGCA-RCKGLPTD 240
 DB 179 RKPLTVLDFASNL-SCKPCHPNCTEDHCWAGEQNCQTLTKVICAQCSGRCKRVPSD 237
 QY 241 CCHECAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMPNPEGRYTFGAS 300
 DB 238 CCHNCAAGCTGPRESDCLACKKFRDDATCKDTCPLVLYNPTTYQMDVNPBGKVSFGAT 297
 QY 301 CVTACPNYLTDSGCTVLCPLHNOVETADGTORCKSKPCARVCYGVGLMEHLREVR 360
 DB 298 CVRECPHNVVTDHSGSVRSNCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGI 356
 QY 361 AVTSANIQEFAGCKIFGLSLFPLPSFDGDPASNTAPLQPEQLQVFTEITGLYLYISA 420

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)
 C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
 C:Accession: A47253
 R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;
 Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
 A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epiderm

DB 357 SINATNIDSFKNCTCKINGDVSLPVAFLGDAFTKTLPLDPKKLDVFTVKEISGFLLIQA 416
 QY 421 WPDLSPLDSVFQNLQVIRGILHNGAYSILTLQGLIGISWLGRLSRLRELGSGLAIHNTHL 480
 DB 417 WPDNATDLVAFENLEIRGTRKHGOYSLAVNVLKIQSLGRLSKISDGDIAIMKNKVL 476
 QY 481 CPVHTVPDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCMGPGOPTCVNCSQF 540
 DB 477 CYADTMNWRSLFATOSQTKIINRNKNKNDADRHYVCDPLCSDVGCMGPGPFHCFSCRF 536
 QY 541 LRQECVEECRVLQGLPREYVNAHCLPCHPECPONG---SVTCFGEADQCACAHYK 597
 DB 537 SRQKECVKOCNIIQEGEPREFRDSKLPCHSECLVQNSTAYNTTCSGPGDHCMKCAFI 596
 QY 598 DPFFCVARCPGKVPDLSPYMPKWFDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPL 657
 DB 597 DGPCHVKACPAVLGENDTL-VWKYADANAVQLCHPNCTRGCKGPGLEGCP---NGSKT 652
 QY 658 TSIVSAV--GILLVVVLGVVFGILIKRRQOKIRKYTMRLQLQETELVEPLTPSGAMPNOA 716
 DB 653 PSIAAGVVGGLLCLLVVGLGIGLYLRR--HIVKRTLRLRLQERLELVEPLTPSGEAPNOA 711
 QY 717 QMRILKETELRKVKVLGSCAFGTVYKGIWIPGENVKIPVAIKVRENTSPKANKEILDE 776
 DB 712 HLRIILKETEFKVKVGLSGAFGTVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDE 771
 QY 777 AYVMAVGSPYVSRLLGICLTSTVOLTMPVGLLDHVNRGRGLSGDOLLNWCQIOA 836
 DB 772 AYVMAVDNPHVCRLLGICLTSTVQITQIMPYGCLLDYIREHKNIGSYLNLNWCQIOA 831
 QY 837 KQMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKQMA 896
 DB 832 KGMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGKVPIKQMA 891
 QY 897 LESILRRRPTHOSDVMVSYGVTVWELMTFGAKYDGIIPAREIDPLLEKGERLPQPPCTID 956
 DB 892 LESILHRIYTHOSDVMVSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPQPPCTID 951
 QY 957 VYIMVVKMIDSECRPRELVSFRSMARDPQRFVVIQ-NEDLGASPLDSTFVRSLL 1015
 DB 952 VYIMVVKMIDADSRPKRSLIAEFSKWARDPPRYLVIOGDERMHLPSPTDSKFRVTL 1011
 QY 1016 EDDNDGDLVDABEYLVPOQGFPCDPAPGAGGVHHRHSSSTRSGGDLTLGLEPSEEE 1075
 DB 1012 EEDMEDIVDADEYLVPHQGF-----NSPST----- 1038
 QY 1076 APRSL-----APSGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSET 1130
 DB 1039 -SRTELLSLSATSNSATNCID-----RNGQGHVPVREDSFVQYSSDPTGNFLEES 1089
 QY 1131 --DGVAPLTCSQPEYVNPQDVRPPSPREGPLPAARAGATLERAKTLPKGKGVK 1188
 DB 1090 IDDGFL-----PAPEYVQ--LMPKPS-----TAMVQNIYN 1120
 QY 1189 DVF-----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQ-- 1231
 DB 1121 NISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNOSPLAKTVFESSPYIQSG 1172
 QY 1232 -----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1260
 DB 1173 NHQINLNDPDYQODFLPNETKPNGLLKVPAENPEYLRVAAP 1214

A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLOC>
A:Cross-references: GB:U07868; NID:g337359; PID:AAB59446.1; PID:g337360
A>Note: sequence extracted from NCBI backbone (NCBI:P126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Query Match 43.4%; Score 2972.5; DB 2; Length 1308;
Best Local Similarity 45.1%; Pred. No. 4.7e-114;
Matches 613; Conservative 185; Mismatches 369; Indels 191; Gaps 31;

QY 9 WGLLLALLPPGAA-----STQVCTGTDMLKRLPASPETHLDMLRHLVYOGQVVOGNLELTY 64
DB 8 WVVSLVLAAGTVQPSDSQVACGTENKSSLDLEQQVRAALRYENCEVVMGNLEITS 67

QY 65 LPTNASLFLQDIQEVQGVYVLIAHNOVRQVPLQRLAIRVGTQLFEDNYALAVLDNGDPLN 124
DB 68 IEHNRDLSELRVREVTGVYVVALNQFVYLPLENLRIIRGTLKYEDRYALAIPLNRYKDG 127

QY 125 NTPVTGASGGURELQRLSLTEILKGGVLIQNPOLCYODTILWKDF-----NNFTVS 178
DB 128 NF-----GLQELGLKXNLTEILNGVVVDQNKFLCYADTIHWQDIVRNWPWSNLT- 177

QY 179 FWLVRVKSVAHLENRSRACHPCSPKCGSRCHGESSEDCOSLTRVTCAGGC-ARCKGPL 237
DB 178 -----VST-----NGSSGCGRCHKCTG-RCMGPTENHCQTLRTVCAEQDCGRCYGY 225

QY 238 PTDCEHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTF 297
DB 226 VSDCHREACGCGSGPKDTCFACMFNDSGACVTCQPTFYVYNTFFQLEHFNKAYTY 295

QY 298 GASCVTACPNYLLSTDVSGCTLVCPHLNQVETHAEDGTQCEKCKSPCARVCYGLGMEHLR 357
DB 286 GAFVCKKCPHNFV-VDSSSCVRACPSSKMEV-BENGIKMKCPKTDICPKACDGIQTGSLM 343

QY 358 EVRAVTSANIOEFAGCKKIFGSLAFPEPSFGDGPASNTAPLOEQLOVFPETLEITGYLY 417
DB 344 SAGTVDSNLIDKFINTKINGNLIFLVTGHHGPDYNAIEAIDPEKLNFRVREITGFLN 403

QY 418 ISAWPDLSPDSVFONQVIRGILHNGAYSLLTLOGLIGISWLGRLSRLGSLGLAIHNN 477
DB 404 IQSWPNNMTDFSVFSLNLTIGRVLVSGLSLLILKQGGITSLOFQSLKEISAGNIYITDN 463

QY 478 THLCFVHTVPWQDLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQVCNC 537
DB 464 SNLCYHTTNWTLTFSTINQRIVIRDNKAENCTAEGMVCNHLICSSDGCWGPQDQCLSC 523

QY 538 SDFLRGOECVEECRVLOGLPREVYNARHCLPCHPECO-ONGSVTCFGEADOCVACAHY 596
DB 524 RFRSRRICIESNLVDGFEFREFENGISVCECDPQCEKMEDEGLTCHGPGPDNCTKCSHF 583

QY 597 KDPFFCVARCPGSKVPLDLYMPKFPFDEEGACQPCPINCTHSCVDLDDKGC----- 648
DB 584 KDGPNVCEKPCDGLQGANSE--IFKYADPDRECHPCPNCTQCGPTSHDCIYYPTGTH 641

QY 649 ----PAQORASPLTSIVSAVV-GILLVVVLGVVFGIILIKRQKIRKYTWRRLLQETELVE 704
DB 642 STLPOHAR-TPL--IAAGVIGGILFVIVGLTFAVYVRRKSIK-KKRALRRFL-ETELVE 696

QY 705 PLTPSGAMPNQAOMRILKETELRKVKVGLSGAGFTYVKGWIIPDGENVKIPVAIKVLRN 764
DB 697 PLTPSGTAPNQAQRLIKETELRKRVKVLGSGAGFTYVKGWIWPEGTEKIPVAIKILNET 756

QY 765 TSPKANKEILDEAYVMAGVSPVSRLLGLTCLTSTVQLVTQMLPQYGLCLLDHVRNRRGLG 824
DB 757 TGPKANVEFMEALIMASMDHPLHVLGLVCLSPITQLVTQMLPMPHCLLLEYVHEHKDNIG 816

QY 825 SQDLLNWCQIAKGMSELYEDRLVLRHDLAARNVVKSPNHVKITDIFGLARLLDIDETEHY 884

DB 817 SOLLNWCQIAKGMSELYEDRLVLRHDLAARNVVKSPNHVKITDIFGLARLLDEGEKEYN 876
QY 885 ADGGKVIKWMALLESILRRFTHQSDVWSYGVVWELMTFGAKPYDGIPIAREIPDLLEKG 944
DB 877 ADGGKVIKWMALLESILRRFTHQSDVWSYGVVWELMTFGAKPYDGIPIAREIPDLLEKG 936
QY 945 ERLPOPPICTIDVYIMVVKCMIIDSECRPRFRELSEFSEFARMARDPQRFVVIQNEE-LGPA 1003
DB 937 ERLPOPPICTIDVYIMVVKCMIIDSECRPRFRELSEFSEFARMARDPQRFVVIQNEE-LGPA 996
QY 1004 SPLDSTFYRSLLDEDDMDGOLVDAEYLVFQQGFFCCDDPAGPAGCMVHRHRSSTSSGCG 1063
DB 997 SPNDSKFFQNLDEEDLEMDAEYLVF-QAFNIPPP-----IYTSRARIIDNSR--- 1046
QY 1064 DLTGLGPESEEAAPS-----PLAP-SEGAGSDVFDG 1094
DB 1047 --EIGHSPPAYTPMSGNOFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQGATAEIFDD 1104
QY 1095 DLGMGAAGLQSLPHTDPSPLQRYSDPTVPLPS-----ETDGYVAPLTCSPQPEYVN 1147
DB 1105 SCNGTGLRKPAPHVQEDSTQRYSDPTVFAPERSPRGELDEEGYTPMRDKPKOEYLN 1164
QY 1148 QPDVRRQPPSPREGPLPAAPAGATLERAKTILSPGKGVYKOVFAFGGAVENPEYITPQG 1207
DB 1165 PVE-----ENPFVSRR-----KNGDLQ-----ALDNPYHNASN 1193
QY 1208 GAAPQHPPPA-----FSPAFDNLVYWDOP 1233
DB 1194 G-----PPKADEYVNEPLYLNTFANTLGKAEYLNKNTLSMPEKAKAPDNPYWNHSL 1247
QY 1234 PERGA--PPSTFKGTPT-----AENPEYL 1255
DB 1248 PRSTLQHPDYLDQYESTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N:Alternate names: epidermal growth factor receptor homolog; kinase-related transform
C:Species: Xiphophorus maculatus (southern platyfish)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C:Accession: S06142; S13809
R:Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Robe
Nature 341, 415-421, 1989
A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu
A:Reference number: S06142; MUID:90015140; PMID:2797166
A:Accession: S06142
A:Molecule type: DNA
A:Residues: 1-1166 <WIT>
A:Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R:Adam, D.; Maeueller, W.; Scharf, M.
Oncogene 6, 73-80, 1991
A:Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphoph
A:Reference number: S13809; MUID:91125882; PMID:1846957
A:Accession: S13809
A>Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A:Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C:Genetics:
A:Gene: mrk
A:Map position: Y
A:Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; t
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 39.0%; Score 2674; DB 1; Length 1166;
Best Local Similarity 45.0%; Pred. No. 6.1e-102;

Db 232 AGCGSPQDTCFACRHFNDGACVPRCPQPLVYNKLTFLQLEPNPHTKYQYGVCVASCP 291
Qy 307 YNYLSTDVSGCSTLVCPPLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 366
Db 292 HNFV-VDTQTCVRACPDQKQVD-RNGLKMCPCGGLCPKACEGTGG--SRFTVDSN 347
Qy 367 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLBEITGYLYISAMPDSL 426
Db 348 IDGFVNCIKILGNLDFLITGLNGDPWHKIPALDPEKLNFTVREITGYLYNQSPPHMH 407
Qy 427 DLSVFONQVIRGRILHNGAYS-LTLOGLGISWLGRLSRLSRELGSGLALIHNTLFCVYT 485
Db 408 NFSVFSNLTTIGRSLYNGFSLIMKNLNVTSLGFRSLKEISAGRIYISANRQLCYHIS 467
Qy 486 VPWDQLFRNPQOALLHTA-NRDEECVGEGLACHOLCARGHCGWPGPTOCVNCOSQLRQ 544
Db 468 LNWTKVLRGFTERLDIKHNRPRDCVAGKVCPLCSSGGCGWPGQCLSCRNSRGG 527
Qy 545 ECVEECRVLOGLPREYVNAHCLPCHPECOQNGSVTCFGEADQCVCAHYKDPFCVA 604
Db 528 VCVTHCNFLNGEPREPAHEAECFSCHEPCQMEGTATCNGSGSDTCAQCAHFRDGPCHVS 587
Qy 605 RCPGSGKPLSYMPIKWFDEGACOPCINCHTSCVDLDDKCCPAEORA-----SPLTSI 560
Db 588 SCPHGVLG--AKGPIYKPDVQNECRPCHENCTQGCKGPELOQCLGQTLVLIGKTLHTWA 645
Qy 661 VSAVVGILLVVLGVVFGILIKRQOKIR-KYTMRELLQETELVEPLTPSGAMPNOAQR 719
Db 646 LTVIAG--LVVIFMMLGGTFLYWRGRIQNKRAMRYLGERGSEIEDFS-EKANVILAR 702
Qy 720 ILKETELRKVKVLGSGAGFTVYKGIWIPDGENYKIPVAIKVIRENTSPKANKEILDEAYV 779
Db 703 IFKETELRKVLGSGVFGTVHKGVWIPGESIKIPVCIKVIEDKSGRSFOAVTDHMLA 762
Qy 780 MAGVGSPPYVSRLLIGLITSTVOLVTQIMPYGCLLDHVRENRGRLSGDILLNWCMTAKGM 839
Db 763 IGSLDHAHIVRLGLCGSLQVLTQYLPGLSLDDHVHRGALGPQLLNMGVQIATGM 822
Qy 840 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPKIMMALES 899
Db 823 YLEEHGMVHRNLAAARNVLKSPSQVOVADFGVADLLPPDDKQLLYSEAKTPIKMALES 882
Qy 900 ILRRRTHQSDVMSYGVTWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYM 959
Db 883 IHFGKYTHQSDVMSYGVTWELMTFCAEPYAGRLAEVDPDLLEKGERLAQPQICTIDVYM 942
Qy 960 IMVKCMIDSECPRERELVSERSMARDPQRFVITQNEIDLGA-----SPLDSTFYRSLLE 1016
Db 943 VMVKCMIDENIRPTEKELANEFTRMARDPPRYLVIKRES-GRGIAPGPEPHGLTNKKLE 1001
Qy 1017 DDDMGDLVDAEYLVPOQGFPCDPAPAGAGMVHHRSSSTRSGGDLTLGLEP-SEEE 1075
Db 1002 EVELEPELDLDDLEAED-----NLATYTLGALSPLVGTILNRRP 1042
Qy 1076 APRSPLAPSEGAGSDVFDGLNGAAKGLQSLPTH-PSPLQRYSEDPTVPLP-----S 1128
Db 1043 GSQSLSPSSGY-MPMNOGLNGBSCSAVSGSSERCPRVSLH-----PMRPGCLASE 1095
Qy 1129 ETDGYA-----PLTCSQPE-----YVNPQVVRPPOPPSPREP----- 1162
Db 1096 SSEGHTVGEAELOEKVSMCRSRSRSPRGRDSAYHSQRHSLLTFTVPLSPGPLEED 1155
Qy 1163 -----LPAARPAGATLERAKTISP-KGNKGV-----KDVFAFGAVENPEYLTPOGGAA 1210
Db 1156 VNGYVMPDTHLKTPTSSRECTLSVGLSVLTGTEEDED-----EVEYNNRRRHS 1207
Qy 1211 PPHPPPPSPADNLIYYWD-----QDPPERGAPSTFTGTPTAENPEYL 1255
Db 1208 P-PHPRPSPSLEELGYEYMDVSGSDLSASLGSTQSCPLHPVPIPTAGTTPDEYEM 1263

RESULT 10
JC4387

epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: J04387
R:Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein
A:Reference number: J04387; MUID:96096535; PMID:8522190
A:Accession: J04387
A:Molecule type: mRNA
A:Residues: 1-1339 <HEL>
A:Cross-references: GB:U29339; NID:G915389; PID:G915390
A:Experimental source: liver
A:Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 370.
C:Comment: This protein is a functional heregulin receptor that transduces signals to C:Genetics:
A:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:840-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)

Query Match 34.2%; Score 2341.5; DB 2; Length 1339;
Best Local Similarity 40.5%; Pred. No. 2.4e-88;
Matches 523; Conservative 174; Mismatches 429; Indels 165; Gaps 35;
Qy 3 LAALCRWGLLLALLPPGAA---STVCTGTCKMLRLPASPEHLDMLRLHYQCGVVOGN 59
Db 7 LQVLC-----FLSLARGSEKNSQAVCPGTGLSLVTDADNQOYQTLKLYKECEVVMGN 62
Qy 60 LELVLPYTNASLFIQIOEVQGYLIAHNVQVPLQRLIRIVRGTLQEDNVALVDN 119
Db 63 LEIVLTGHNADLSFLQWIREVAYVLVAMNFEVSLPLNLRVVRGTQVVDGKFAIFVM-- 120
Qy 120 GDPANNTTPTVGASPGLELQRLSLTEILKGVLIQRPOLCYOTILWKDFNNFTVSF 179
Db 121 ---LNTYNT---NSSHALRQLKFTQLEITLSGVYIEKNDKLCMDTIDMRD----- 165
Qy 180 WLRVPKVSASHL--ENRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGC- ARCKGP 236
Db 166 ---IVRVRAEIVKNGANGCPCHEVCKG-RCWGPDPDCCQLLTITICAPQCNCRGCP 221
Qy 237 LPTDCHEQCAAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 222 NPNQCCHDECAGCGSPQDTCFACRRFNDSGACVPRCPPLVYNKLTFLQLEPNPHTKYQ 281
Qy 297 FGASCVTACPNYVLTVDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHL 356
Db 282 YGVVCVASCPHNFV-VDTQTCVRACPDQKQVD-RNGLKMCPCGGLCPKACEGTGG-- 337
Qy 357 REVRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLBEITGYL 416
Db 338 SRYQTVDSNIDGFVNCIKILGNLDFLITGLNGDPWHKIPALDPEKLNFTVREITGYL 397
Qy 417 YTSAMPDLSPLDSVFONQVIRGRILHNGAYS-LTLOGLGISWLGRLSRLSRELGSGLALIH 475
Db 398 NIQSWPPHMHNFVFSNLTTIGRSLYNGFSLIMKNLNVTSLGFRSLKEISAGRIYIS 457
Qy 476 HNTHLCFVHTVWDQLFRNPQOALLHTA-NRDEECVGEGLACHOLCARGHCGWPGPTQC 534
Db 458 ANQOQLCYHHSLNWTLRLRGPSERLDIKYDRPLGECLAEQKVCPLCSCGCGWGPAGQC 517
Qy 535 VNCOSPLRQCEVEECRVLOGLPREYVNAHCLPCHPECOQNGSVTCFGEADQCVCA 594
Db 518 LSCRNYSRGVCVTHCNFLQGEPRFVHEAQCFSCHEPCLPMEGTITNGSGSDACARCA 577
Qy 595 HYKDPFPFCVARCPSGVKPLDSYMPIKWFDEGACOPCINCHTSC--VDLDDKGPAP 652

Db 578 HFRDGPCHVNSCPHGILG--AKGPIYKYPPDAQNECRPCHECTQCNGPELQDCIGQAEV 635
Qy 653 RASPLTSIVSAVGVLLVWVGVGILIKRRQOKIR-KYTMRRLLQETELVEPLTPSGA 711
Db 636 LMSRPHLVIATVG--LAVILMILGSGFLYWRGRIOKRAMRRYLERGESIEPLDPS-E 692
Qy 712 MPNQOMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 771
Db 693 KANKVLARIFKETELRKVLKSGVFTVHKGIWIPGEGSIKIPVICIKVIEDKSGRQSFQ 752
Qy 772 EILDEAYVMAGVSGPYVSRLLGICLTSTVOLVTLMPYGCCLLDHVRENRLGSDQLNW 831
Db 753 AVTDHMLAVGSLDHAHIVRLGLCPGSSQLVTVPLGSLLDHVQKHRETLPQQLLNW 812
Qy 832 CMQIAKMSYLEDVRLHVRDLAARNVLPKSNHVKITDFGLARLLDIDETEHADGKGVP 891
Db 813 GVQIAKMGYYLEHSMVHRDLARNVLMKSPSQVQVADGVADLLPPDDKQLLHSEAKTP 872
Qy 892 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLP 951
Db 873 IKWMALESIFHGKYTHQSDVWSYGVTVWELMTFGAEPYAGRLABIPDLEKGERLAQ 932
Qy 952 ICTIDVYMWKCMIDSECRPRPRELVSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFY 1011
Db 933 ICTIDVYMWKCMIDENIRPTFKELANEFTRMARDPPLYLVIKRAS-GPGTP--PAAE 989
Qy 1012 RSLDEDDMDGLVDAEYLVPOQGFPCPDPAFGAGGVHRRHRSSTRSGGDLTLGLEP 1071
Db 990 PSVLTTEL-----QEAELEPEL-----DLDLLEA 1015
Qy 1072 SEE-----EAPRSLPASEG-----AGSDVDFDGLGMAAGK 1103
Db 1016 EEEGLATSLGALSUPTGTLTRPGQSLLSPSSGYMPMNOSSSLGEACLDASVAGRG 1075
Qy 1104 LQSLPHTDPSPLQRYSEDDTVPLPSETDGV-----APL-----TC-----SPOE 1144
Db 1076 SRPISLH-PIPRGR-----PASESSEGHVTCSEAELOEKVSVCSRSPRGRDGS 1127
Qy 1145 -YVNPQDVRPPOPPSPREGP-----LPAARPAGATLERAKTLP-SRNGVYV----- 1187
Db 1128 AYHSQRHSLLTPTVPLSPGLLEEDGNGVMPDTHLRGASSSREGTLLSGVLSVLGTEE 1187
Qy 1188 --KQVFARCGAVENPEYLTPOGGAAPQPHPP 1216
Db 1188 EDED EDED-----EEYEYMNKRGRGSP-PRPP 1209

RESULT 11
TVFVLV
Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and product formation
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <NIL>
A:Cross-references: GB:M10066; GB:M13801; NID:g211749; PIDN:AAA48763.1; PID:g211750
A:Note: In Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gallus
C:Comment: This protein is synthesized as a gag-env-erbB protein.
A:Gene: gag-env-erbB
A:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific protein kinase
F:1-6/Product: gag protein (fragment) #status predicted <GAG>
F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F:194-459/Domain: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif

F:229/Active site: Lys #status predicted

Query Match 25.8%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred No. 3.4e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;
Qy 584 GPEADQCACAHYKDPFFCVARCPGKPDLSYMPIMKFPDEBEGACQPCPINCCHSCVDL 643
Db 60 GP--DHCMKAFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHPNCNTRCKGP 116
Qy 644 DDGCPAPQASPLTSIVSAVY-GILLVVLGVFGILIKRRQOKIRKVTMRRLLOETEL 702
Db 117 GLEGCP--NGSKTPIAAGVVGGLCLVWVGLGILGLYLR--HIVRKTRRLLOEREL 172
Qy 703 VEPLTPSGAMPNQAMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLR 762
Db 173 VEPLTPSGEAPNQAHRLILKETEFKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKELR 232
Qy 763 ENTPSKANKEILDEAYVMAGVSGPYVSRLLGICLTSTVOLVTLMPYGCCLLDHVRENRR 822
Db 233 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLVTLMPYGCCLLDVIREHKDN 292
Qy 823 LGSODLWKCQIAKMSYLEDVRLVHRDLAARNVLPKSNHVKITDFGLARLLDIDETE 882
Db 293 IGSQYLLWKCQIAKGMNLEERLVRDLAARNVLPKTPQHVKITDFGLAKLLGADEKE 352
Qy 883 YHADGKGPYIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIP 942
Db 353 YHAEKGKVPYIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASEISSVLE 412
Qy 943 KGERLPQPICTIDVYMWKCMIDSECRPRPRELVSEFSRMARDPQRFVVIQ-NEDLG 1001
Db 413 KGERLPQPICTIDVYMWKCMIDADSRRPKFRELIAEFSQWARDPPRYLVVIQGERMH 472
Qy 1002 PASPLDSTFYSLLEDDMDGLVDAEYLVPOQGFPCPDPAFGAGGVHRRHRSSTRSG 1061
Db 473 LPSPTDSKYRTLMEEDMEDIVDADEVLPVHQGF-----NSPST--- 513
Qy 1062 GGDLTGLGLEPSEEAAPRSP-----APSEGAGSDVDFDGLGMAAGKLSLTHDPSPLO 1116
Db 514 -----SRTPLSSLSATSNNSATNCID-----RNGQHPVRESFVQ 550
Qy 1117 RYSEDPTVPLPSET--DGYVAPLTCSPQPEYVQPDVRPQPSREGPLPAARPAGATLE 1174
Db 551 RYSSDPTGNFLEESIDDGFL-----PAPEYVQ--LMPKPS----- 585
Qy 1175 RAKTLSPKNGVGVKQVF-----AFGAVENPEYLTPOGGAAPQPHPPAP 1219
Db 586 ----TAMVQNCIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLA 633
Qy 1220 SPADNLYYWDQ-----DPPE-----RGAPPSTFGTPTAENPEYVLGLDVP 1260
Db 634 KTVFESSPYWIQSGNHQINLNDPDIYQDDFLPNETKPNGLLKVPAAENPEYLRVAAP 689

RESULT 12
TVYUWH
Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C>Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A:Reference number: A00644; MUID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R:Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Marti
Science 224, 1456-1459, 1984
A:Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type
A:Reference number: A38022; MUID:84223957; PMID:6328658

A;Reference number: S35743
A;Accession: S35745
A;Molecule type: DNA
A;Residues: 1-544 <VEN>
A;Cross-references: EMBL:X
C;Genetics:

C;Genetics:
A:Gene: *erbB*
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

Query Match 24.1%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 2e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

| | Matches | 343; | Conservative | 70; | Mismatches | 121; | Index | 92; | Gaps | 15 | | |
|----|---------|---------|--------------|--------|------------|------|---------|-------|--------|------|------------|-----|
| QY | 584 | GPEADQC | VACAHYK | DPFFCV | ARCP | SGYK | PDLSYMP | IKWFP | DEEGAC | QPCP | INCTHSCVDL | 643 |

| | | | |
|----|-----|--|-----|
| Db | 1 | GP--DHCKAHFIDGPHCVKACPAAGLVGENDTL-VVKYADANAVCQLCHPNCTRGGKP | 57 |
| QV | 644 | DDKGCPAEQRASPLTSIVSAVV-GILLVVVLGWVFGLIKRRQQKIRKYTMRLRLLQETEL | 702 |

Db 58 GLEGCP---NGSKTPSIAAGVVGGLCLVWGLGIGLYLRRR-HIVKRTLRLLQEREL 113

| | Qy | Db |
|-----|---|-----|
| 703 | VEPLTPSGAMPNQAMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLR | 762 |
| 114 | VEPLTPSGEAPNQAHRLKETETFKVKVLGFGAFGTVYKGIWIPGEKVPVAIKELR | 173 |

QY 763 ENTSPKANKEIILDEAYVMAGVCSPPVSRLLGICLTSTVQVLTQTMPCYCLLDHVRENRR 822

| | | | |
|----|-----|--|-----|
| DB | 174 | EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDN | 233 |
| QY | 823 | LGSDLLNLCMGOIAGKWSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE | 882 |

| | | | |
|----|-----|--|-----|
| DB | 234 | IGSQYLLNNVCVOIAKGMNLYEERHMHVDRDLAARNVLKTPQHVKYITDFGLAKQLGADEKE | 293 |
| QV | 883 | YHADGGKVPVIAKMALESTLRREFETHOSDWSVSGVTVMFLMTFCACKPDGIPAREIPDILE | 942 |

[illegible]

| | | | |
|----|-----|--|------|
| Qy | 943 | KGRLPOPPICITIDVYIMVVKCWMIDSECRPRFRELSEFSEFMRDQRFVVIQ-NEDLG | 1001 |
| Db | 354 | KGRLPOPPICITIDVYIMVVKCWMDSRPFRELIAEFSEKWARDPPRYVITQGDERMH | 413 |

QY 1002 PASPLDSTFYRSLLEDDMGDLVDAEYLVPOQFFCFDDPAPCAGGVMVHRRHSSSTRSG 1061

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|------|------|----|----|----|----|----|----|---|---|---|---|---|---|---|---|-------|----|---|---|---|---|-------|----|---|---|---|-----|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|------|
| Db | 414 | LRSP | TD | SK | FY | RL | ME | ED | M | D | V | A | D | E | V | L | P | H | Q | G | F | F | ----- | NS | P | S | T | --- | 454 | | | | | | | | | | | | | | | | | | | | | |
| Qy | 1062 | GGD | L | T | L | G | L | P | S | E | E | A | P | R | S | L | ----- | AP | S | E | G | A | G | S | D | V | F | G | D | L | G | M | C | A | A | K | L | Q | S | I | P | T | H | D | P | S | P | L | Q | 1116 |

Db 455 -----SRTPLLSLSATSNNSATNCIDRNGG-----H-----481

Dz

Qy Qy Dz

1117 RYSEDPTVPLPSETDGVAPLTCSPQEYVNQPVDVRPOPPSPREGPLPAARPAGAT--LER 117
:::
|||
482 -----PVREDGEL-----PAYEYNQ--LMPKKPSAMVQNQIYYNYSILTAISK 523

| | | | |
|----|------|----------------------------|------|
| QY | 1176 | AKTSPKNGVVKDVFAGGAVENPEYL | 1203 |
| | | | : |
| Dh | 524 | LDTSNBYCN-----SUSTAVDNPEYL | 544 |

DD 524 LFIDSRIGN-----SHSTAVDNPEYL 544

RESULT 14

epidermal growth factor receptor - fruit fly (*Drosophila melanogaster*)
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: *Drosophila melanogaster*
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C:Accession: A00640; A38021

Rilivneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A title: The Drosophila EGF receptor gene homolog: conservation of both hormone

A,Reference number: A00640; MUID:85124611; PMID:2982499

Qy 511 --VOEGLA-----CHQLCARGHCWGPGTQCVCNCSQFLRGQECVCECRVLQGLPREYV 561
Db :|||:|||:
Qy 597 AGIAEPLAGRAVCKKCHPLCELCTNYGYHEQVCKCTHYKRQOCETEC-----PADHY 650
Db :|||:|||:
Qy 562 ---NARHCLPCHPECQPONGSVTCFGEADQCACAHYK-----DPFF-----CVAR 605
Db :|||:|||:
Qy 651 TDEQRCECFORHEPC---NG---CTGPGADDCKSCRNFKLFDANETGPYNVNSTMFNCSTSK 704
Db :|||:|||:
Qy 606 CPSSGVK-PDLSYMPIMKFPDEEGACQPCPNCTHSCVDLDKGCFAQERASPLTSIVSVA 664
Db :|||:|||:
Qy 705 CPLEMRHVNTQYTAIGPY-----CAASPRRSKITANLD-----VNMIIFI 745
Db :|||:|||:
Qy 665 VGIILVVVLGWVFGEI-LIKRROQKIRRYT--MERLLQETELVELBPLTPSGAMPNOAQRIL 721
Db :|||:|||:
Qy 746 TGAVLVPTICILCVTVTYCQKQKAKKETVMKTALMSGRESEPLRPNSNIGANLCKLRIV 805
Db :|||:|||:
Qy 722 KETELRKVKVLGSGAFGTYYKGWIIDGENVKIPVAIKVLRENTSPKANKEIIDEAAYMA 781
Db :|||:|||:
Qy 806 KDAELRKGGVYLHGAFGRVYKGVWVPEGENVKIPVAIKELLKSTGAESSEEFLEAREYIMA 865
Db :|||:|||:
Qy 782 GVGSPPVSRLLGICLTSTVOLVTOLMYPYCGLLDHVRENRCRLGSDLLNMCIOAKMSY 841
Db :|||:|||:
Qy 866 SEBHVNLKLULAVCMGSOOMLIITOLMPLGCLLDYVRNRDRDKIGSKALLNNSTOIAKMYS 925
Db :|||:|||:
Qy 842 LEDVRLVHRDLAAARNVLVK---SPNVKVIDTFGLARLLDIDETEHYADGKVPIKNWALE 898
Db :|||:|||:
Qy 926 LEEKRLVHRDLAARNVLRLLAGEDH----DFGLAKLLSSDSNEYKAAGCKMPIKWLALE 981
Db :|||:|||:
Qy 899 SIURRPTHOSQWSVGTVWELMTFGAPYDGIIPAREIPDLLEKGERLPQPPICTIDVY 958
Db :|||:|||:
Qy 982 CIRNVFTSKSDVWAFGVTTIWELLTFGORPHENIPAKDIPDLIEVGLKLEQPEICSLDIY 1041
Db :|||:|||:
Qy 959 MIWVKWMIDSECRPFRELVSFEFSRMARDPQRFVIONEDLG--PASPLDSTFYRSLLE 1016
Db :|||:|||:
Qy 1042 CTLLSCHWLDDAARMPTFKOLTITVFAEPADPGRYTLAILGDKFTRLPA-----YTSQE 1094
Db :|||:|||:
Qy 1017 DD---DMGDLVDAAEYLYVPQQGFCDPAPGAGGVVHHRRHSSTRSGGDLTLGLEPSE 1073
Db :|||:|||:
Qy 1095 KDLIRKLAPTDOGEAIARPDDVLOPKAALGPS-----HRTDCT----- 1133
Db :|||:|||:
Qy 1074 EAP- ----RSPLAPSEGAGSDVFDG---DLGMGAAGLGLSLTHDPSPLOYSEDPTV 1124
Db :|||:|||:
Qy 1134 DEMPKLNRYCKOPSNKNSSTGDDERDSSAREVEGVGNLR-----L 1172
Db :|||:|||:
Qy 1125 PLPSETDGVVAPLITCSPOEVYVNPQDVROPSPREGPLPAARPACATLERAKTLSPGXN 1184
Db :|||:|||:
Qy 1173 DLVPDEDDYMLP-TCQPGNNNNMMN-----NPNQNMAAVGAAGYM----- 1214
Db :|||:|||:
Qy 1185 GVVKYDVFAGGAVENPEYL---TPQGGAAPQH-----PPPAFSP- 1221
Db :|||:|||:
Qy 1215 ----DLIGVPVSDNPEYLLNAQTGLGVGESPIQTOTIGIPVMGGPGTMVKVPMPCSEPT 1270
Db :|||:|||:
Qy 1222 AFDNLYYWD 1230
Db :|||:|||:
Qy 1271 SSDHEYND 1279

RESULT 15
S00727
Kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C;Accession: S00727
R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus
A;Reference number: S00727; UID:88217326; PMID:2897102
A;Accession: S00727
A;Molecule type: DNA
A;Residues: 1-545 <SCO>
A;Cross-references: EMBL:X06943
C;Genetics:
A;Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

```
Query Match      23.9%; Score 1640; DB 2; Length 545;
Best Local Similarity 54.9%; Pred. No. 3.8e-60;
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

QY 584 GPEADQCACAHYKPPFCVPCVARGVGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDL 643
DB 1 GP--DHCCKCAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCTRGCKGP 57

QY 644 DDKGCPAEORASPLTSIVSAW-GILLVVVLGVVFGILIKRQOKIRKYTMRELLQETEL 702
DB 58 GLEGCP---NGSKTPSIAGVGGLLCLVVGIGLGLYLR--HIVKRTLRLLQEREL 113

QY 703 VEPLTSGAMPNCAQMRILKETELRKVKVLGSGAFGVYKGIWIPGENVKIPIVAIKVL 762
DB 114 VEPLTSGEAPNQAHLRIKETEFKKVKVLGFGAFGVYKGLWIPEGEKVTIPIVAIKEL 173

QY 763 ENTSPKANKEIIDEAYVMAGVSGPYVSRLLIGICLTSTVQLVTOLMPYGCLLDHVRENRR 822
DB 174 EATSPKANKEIIDEAYVMASVDNPHVCRLLGICLTSTVQLITOLMPYGCLLDIYREHKON 233

QY 823 LGSQDLLNMCMIKAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
DB 234 IGSQYLLNWCQIAKGMNLEERHLVHRDLAARNVLVKTPODVKITDFGLAKQLGADKE 293

QY 883 YHADGCKVPIKMALESILRRRTHQSDVMSYGVTVWELMTFGAKYDGIPIAREIPDLLE 942
DB 294 YHAEGGKVPIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPASEISVLE 353

QY 943 KGERLPQPPICTIDVTMIMVKMIDSECRPRRELVSFSPRMARDPQRFVVIQ-NEDLG 1001
DB 354 KGERLPQPPICTIDVTMIMVKMWDSDSPKPRELIAEFKWARDPPRYLVLIQDERMH 413

QY 1002 PASPLDSTFYRSLLDDMGDLVDABEYLVPOQGFPCPDPAAGAGMVRHRRSSSTRSG 1061
DB 414 LPSPTDSKFYRTLMEEDMEDIVDADAYLVPHQGF-----NSPST--- 454

QY 1062 GGDLTGLRPSSEEA PRSPL-----APSEGAGSDVFDGDLGMAKGLQLSLPTHPSPLQ 1116
DB 455 -----SRTPLLSLSATSNNSATNCIDRNG-----H----- 481

QY 1117 RYSEDPTVPLPSETDGYVAPLTCSPQPEYVNPQDVRPQPSPREGPLPAARPAAGAT-LER 1175
DB 482 -----PVREDGFL-----PAPEYVQ--LMPKPESTAMVQNIYNYISLTAISK 523

QY 1176 AKTLSPGKNGVVKDVFAGGAVENPEYL 1203
DB 524 LPMDSRYQN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:08:44
Job time : 32.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-149-163-14
Perfect score: 6848
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES | | | | |
|------------|--------|-------------|--------------|---|
| Result No. | Score | Query Match | Length DB ID | Description |
| 1 | 6709 | 98.0 | 1255 | P04626 homo sapien |
| 2 | 5924 | 86.5 | 1257 | P06494 rattus norv |
| 3 | 5914.5 | 86.4 | 1254 | ERB2_RAT |
| 4 | 3162 | 46.2 | 1210 | P00553 mesocricetu |
| 5 | 3142 | 45.9 | 1210 | P00533 homo sapien |
| 6 | 2972.5 | 43.4 | 1308 | Q01279 mus musculu |
| 7 | 2959 | 43.2 | 1308 | Q15303 homo sapien |
| 8 | 2690.5 | 39.3 | 1167 | Q22956 rattus norv |
| 9 | 2427.5 | 35.4 | 1342 | P13388 xiphophorus |
| 10 | 2364.5 | 34.5 | 1339 | P21860 homo sapien |
| 11 | 1964 | 28.7 | 1426 | Q62799 rattus norv |
| 12 | 1749.5 | 25.5 | 634 | P04412 drosophila |
| 13 | 1703 | 24.9 | 604 | P00534 avian leuko |
| 14 | 1630 | 23.8 | 540 | P00535 avian eryth |
| 15 | 1611 | 23.5 | 703 | P11273 avian eryth |
| 16 | 1302 | 19.0 | 1323 | P13387 gallus gall |
| 17 | 1142.5 | 16.7 | 245 | P24348 caenorhabdi |
| 18 | 734 | 10.7 | 1363 | P70424 mus musculu |
| 19 | 714 | 10.4 | 1297 | Q02466 branchiosto |
| 20 | 714 | 10.4 | 1300 | P14616 homo sapien |
| 21 | 707 | 10.3 | 1382 | Q9wt14 mus musculu |
| 22 | 699.5 | 10.2 | 1300 | P06213 homo sapien |
| 23 | 698 | 10.2 | 1383 | P14617 cavia porce |
| 24 | 697 | 10.2 | 1607 | P15127 rattus norv |
| 25 | 696.5 | 10.2 | 1372 | Q25410 lymnaea sta |
| 26 | 665 | 9.7 | 1477 | P25208 mus musculu |
| 27 | 649 | 9.5 | 1367 | Q25197 hydra atten |
| 28 | 640 | 9.3 | 1373 | P08069 homo sapien |
| 29 | 636.5 | 9.3 | 1370 | Q60751 mus musculu |
| 30 | 622 | 9.1 | 1390 | P24082 rattus norv |
| 31 | 607 | 8.9 | 987 | Q93105 aedes aegypt |
| 32 | 591.5 | 8.6 | 984 | P54760 homo sapien |
| 33 | 591 | 8.6 | 2146 | Q07494 gallus gall P09208 drosophila |

| | | | | | | |
|----|-------|-----|------|---|------------|--------------------|
| 34 | 589.5 | 8.6 | 977 | 1 | EPA2_MOUSE | Q03145 mus musculu |
| 35 | 588 | 8.6 | 1114 | 1 | RET_HUMAN | P07949 homo sapien |
| 36 | 585 | 8.5 | 987 | 1 | EPB4_MOUSE | P54761 mus musculu |
| 37 | 584.5 | 8.5 | 976 | 1 | EPA2_HUMAN | P29317 homo sapien |
| 38 | 577.5 | 8.4 | 985 | 1 | EPBA_XENLA | Q91571 xenopus lae |
| 39 | 574.5 | 8.4 | 984 | 1 | EPB1_RAT | P09759 rattus norv |
| 40 | 573.5 | 8.4 | 1053 | 1 | FAK1_CHICK | Q00944 gallus gall |
| 41 | 570 | 8.3 | 902 | 1 | EPBB_XENLA | Q91736 xenopus lae |
| 42 | 569 | 8.3 | 1068 | 1 | FAK1_XENLA | Q91738 xenopus lae |
| 43 | 568.5 | 8.3 | 984 | 1 | EPB1_HUMAN | P54762 homo sapien |
| 44 | 567 | 8.3 | 757 | 1 | HT16_HYDAT | P53356 hydra atten |
| 45 | 563 | 8.2 | 1052 | 1 | FAK1_MOUSE | P34152 mus musculu |

ALIGNMENTS

RESULT 1

ERB2_HUMAN

ID ERB2_HUMAN STANDARD; PRT; 1255 AA.

AC P04626;

DT 13-AUG-1987 (Rel. 05, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)

DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell

DE surface receptor HER2) (MLN 19).

GN ERBB2 OR HER2 OR NGL OR NEU.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86118663; PubMed=3003577;

RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,

RA Saito T., Toyoshima K.;

RT "Similarity of protein encoded by the human c-erb-B-2 gene to

RT epidermal growth factor receptor.";

RL Nature 319:230-234(1986).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=86070181; PubMed=2999974;

RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,

| RA | McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J., | |
| RA | Frankie U., Levinson A., Ullrich A.; | |
| RT | "Tyrosine kinase receptor with extensive homology to EGF receptor | |
| RT | shares chromosomal location with neu oncogene."; | |
| RL | Science 230:1132-1139(1985). | |
| RN [3] | | |
| RP | SEQUENCE OF 737-1031 FROM N.A. | |
| RX | MEDLINE=86016729; PubMed=2995967; | |
| RA | Semba K., Kamata N., Toyoshima K., Yamamoto T.; | |
| RT | "A v-erbB-related protooncogene, c-erbB-2, is distinct from the | |
| RT | c-erbB-1/epidermal growth factor-receptor gene and is amplified in a | |
| RT | human salivary gland adenocarcinoma."; | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985). | |
| RN [4] | | |
| RP | VARIANTS VAL-654 AND VAL-655. | |
| RX | MEDLINE=93194196; PubMed=8095488; | |
| RA | Ehsani A., Low J., Wallace R.B., Wu A.M.; | |
| RT | "Characterization of a new allele of the human ERBB2 gene by allele- | |
| RT | specific competition hybridization."; | |
| RL | Genomics 15:426-429(1993). | |
| CC -1- | FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, | |
| CC | ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A | |
| CC | POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF- | |
| CC | ALPHA AND AMPHIREGULIN. | |
| CC | -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein | |
| CC | tyrosine phosphate. | |
| CC | -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS | |
| CC | (POTENTIAL). | |
| CC | -1- SUBCELLULAR LOCATION: Type I membrane protein. | |


```

OY 717 QMRILKTELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETILDE 776
DB 711 QMRILKTELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETILDE 770
OY 777 AYVMAGVGSYPVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSODLLNWCQIA 836
DB 771 AYVMAGVGSYPVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSODLLNWCQIA 830
OY 837 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYADGGKVPKMA 896
DB 831 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYADGGKVPKMA 890
OY 897 LESILRRRPHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTID 956
DB 891 LESILRRRPHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTID 950
OY 957 VYIMVWKCMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
DB 951 VYIMVWKCMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010
OY 1017 DDMGDLVDAEYLVPOQGFCDPAPAGCMVHRRHRSSTSGGGDLTLGLEPSEEA 1076
DB 1011 DDMGDLVDAEYLVPOQGFCDPAPAGCMVHRRHRSSTSGGGDLTLGLEPSEEA 1070
OY 1077 PRSPLAPSEGAGSDVFDGDLGMAKGLSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1136
DB 1071 PRSPLAPSEGAGSDVFDGDLGMAKGLSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1130
OY 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKVDFAFGGA 1196
DB 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKVDFAFGGA 1190
OY 1197 VENPEYLTPOGGAAPQHPPPAPSPAFDNLVYWDQPPPERGAPSTFKGTPTAENPEYLG 1256
DB 1191 VENPEYLTPOGGAAPQHPPPAPSPAFDNLVYWDQPPPERGAPSTFKGTPTAENPEYLG 1250
OY 1257 LDVPV 1261
DB 1251 LDVPV 1255

RESULT 2
ERB2_RAT STANDARD; PRT; 1257 AA.
AC F06494;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (epidermal growth factor
DE receptor-related protein).
GN ERB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=86118662; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The new oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230(1986).
[2]
SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).

```

```

[3]
RN STRUCTURE BY NMR OF 650-668.
RP MEDLINE=92115181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RT oncogenic and oncogenic forms of the neu protein.";
RL EMBO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X03362; CAA27059.1; ALT_INIT.
CC PIR: A24562; TVRTNU.
CC HSP: P11362; IFGK.
CC InterPro: IPR000494; EGFR_L domain.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_pkinase.
CC InterPro: IPR004019; YLP_motif.
CC Pfam: PF00069; pkinase; 1.
CC Pfam: PF00757; Furin-like; 1.
CC Pfam: PF01030; Recep_L domain; 2.
CC Pfam: PF02757; YLP_2; 1.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART: SM00261; FU; 3.
CC SMART: SM00219; TyrKC; 1.
CC PROSITE: PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE: PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE: PS00111; PROTEIN KINASE DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 677 POTENTIAL.
FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 369 CVS-RICH.
FT DOMAIN 473 646 CVS-RICH.
FT DOMAIN 722 989 PROTEIN KINASE.
FT NP_BIND 728 736 ATP (BY SIMILARITY).
FT BINDING 755 755 ATP (BY SIMILARITY).
FT ACT_SITE 847 847 BY SIMILARITY.
FT DISULFID 196 205 BY SIMILARITY.
FT DISULFID 200 213 BY SIMILARITY.
FT DISULFID 221 228 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 312 BY SIMILARITY.
FT DISULFID 316 332 BY SIMILARITY.
FT DISULFID 335 339 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.

```

```
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.5%; Score 5924; DB 1; Length 1257;
Best Local Similarity 86.7%; Pred. No. 3.3e-308;
Matches 1097; Conservative 51; Mismatches 104; Indels 14; Gaps 5;

Qy 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPEHLDMLRHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGIAGTQCTGDMKRLPASPEHLDMLRHLYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIHANQVQVPLQRLRVGTQFQEDYALAVLDNG 120
Dy 61 ELTYVPANASLSFLQDIOEVQGYVLIHANQVQVPLQRLRVGTQFQEDYALAVLDNR 120
Qy 121 DPLNNTTTPVT-GASPGGLRELQLRSLTEILKGGVLIQORNPOLCYQDTILWKDF---NNFT 176
Dy 121 DPQDNVAASTPGRTEGLELRLQLRSLTEILKGGVLIQORNPOLCYQDNVWLKDVFRKN-- 178
Qy 177 VSWLRVPKVSASHLENRACHPCSPMKXGSRGWSSESDCQSLTRTVACAGCARCKGP 236
Dy 179 -----QLAPVDID--TNRACPPCAPACKDNHCWGESPEDCQILTGTICTSGCARCKGR 231
Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYT 296
Dy 232 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYT 291
Qy 297 FGASCVTACPNYLSLTDVSGCTLVCPHLNQVTAEDGTQRCCKSKPCARVCYGLGMEHL 356
Dy 292 FGASCVTTCPNYLSLTVSGCTLVCPNNQVTAEDGTQRCCKSKPCARVCYGLGMEHL 351
Qy 357 REVRAVTSANIQEPAGCKKIPGSLAFLPESPDGDPASNTAPLQEPQLQVFTLEETGYL 416
Dy 352 RGARAITSDNVQEPDGCKKIFGSLAFLPESPDGDPSSGIAPLRPEQLQVFTLEETGYL 411
Qy 417 YISAWPDLPLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLSLRELGSGLALIH 476
Dy 412 YISAWPDLRLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLSLRELGSGLALIH 471
Qy 477 NTHLCFVHTVPWDLFRNPHQALLHTANRPEDE-CVGEGLACHOLCARGHCWGPGTQCV 535
Dy 472 NAHLCFVHTVPWDLFRNPHQALLHNSGRNPEDELCVSSGLVNSLCAHCHWGPGTQCV 531
Qy 536 NCSOFLRGQECVECRVLQGLPRYVNAHCLCPHCECPQNGSVTCFPGPEADQVCAH 595
Dy 532 NCSHFLRGQECVECRVWKGPLPREYVSDKRLCPHCECPQNSSETCFGSEADQCAAH 591
Qy 596 YKDPFFCVARPCSKVDPLSWMPYLPKPDDEGACPCPINCTHSCVDLDKGCAPORAS 655
Dy 592 YKDSSCVVARPCSKVDPLSWMPYLPKPDDEGIQPCPCINCTHSCVDLDKGCAPORAS 651
Qy 656 PLTSIVSAVGLILVWVVGFLIKRRQKIRKYMRLQLQETELVEPLTPSGAMPNQ 715
Dy 652 PVTFIATVWGLLFLILVWVGILIKRRQKIRKYMRLQLQETELVEPLTPSGAMPNQ 711
Qy 716 AQMRILKETELRKVKVLGSGAFGTGVIWIPDGENVKIPVAIKVLRNTPSPKANKBILD 775
```

```
Db 712 AQMRILKETELRKVKVLGSGAFGTGVIWIPDGENVKIPVAIKVLRNTPSPKANKBILD 771
Qy 776 EAYVMAGVSPYVSRLLGICLTSTVQLVTQMLPQYGCCLLDHVRNRRGLSGODLLNWCQI 835
Dy 772 EAYVMAGVSPYVSRLLGICLTSTVQLVTQMLPQYGCCLLDHVRHRRGLSGODLLNWCQI 831
Qy 836 AKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGKGVPIKWM 895
Dy 832 AKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGKGVPIKWM 891
Qy 896 ALSEILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPARIPDLLEKGERLPQPICTI 955
Dy 892 ALSEILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPARIPDLLEKGERLPQPICTI 951
Qy 956 DVYMIWVKCMIDSECPRELVSEFSRMDRQRFVWIONEDLGPASPLDSTFYRSL 1015
Dy 952 DVYMIWVKCMIDSECPRELVSEFSRMDRQRFVWIONEDLGPSSPMDSFYRSL 1011
Qy 1016 EDDMDGLVDABEYLVPOQGFCCPDPAAGGVMVHRRSSSTRSGGDLTLGLEPSEEE 1075
Dy 1012 EDDMDGLVDABEYLVPOQGFCCPDPAAGGVMVHRRSSSTRSGGDLTLGLEPSEEG 1071
Qy 1076 APRSPLAPSEAGSDVFDGLGMGAAGKLSLTPHDSPLQRYSEDPVPLPSETDGYVA 1135
Dy 1072 PPRSPLAPSEAGSDVFDGLGMGAAGKLSLTPHDSPLQRYSEDPVPLPSETDGYVA 1131
Qy 1136 PLTCSPOPEYVQDVPQPPSPREGPLPAARPAAGATLAKTLSPGKNGVVKDVFAGG 1195
Dy 1132 PLACSPQPEYVQDVPQPPSPREGPLPAARPAAGATLAKTLSPGKNGVVKDVFAGG 1191
Qy 1196 AVENPEYLTTPQGAAPQHPHPPAFSPAFDNLVYWDQDPPERPAPPTFKGTPTAENPEYL 1255
Dy 1192 AVENPEYLVREGTASPHPSPAFSPAFDNLVYWDQDPPERPAPPTFKGTPTAENPEYL 1251
Qy 1256 GLDVPV 1261
Dy 1252 GLDVPV 1257

RESULT 3
ID ERB2_MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-253(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
```

CC RESIDUES.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D16295; BAA03801.1; -.

DR HSSP; P11362; 1FCG.

DR InterPro; IPR000494; EGFR_L_domain.

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001245; Tyr_pkinase.

DR InterPro; IPR004019; YLP_motif.

DR Pfam; PF00069; pkinase; 1.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF02757; YLP; 2.

DR ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00219; TyKc; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.

DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;

KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

KW Proto-oncogene; Disease mutation.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.

FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 653 675 POTENTIAL.

FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 158 368 CYS-RICH.

FT DOMAIN 472 644 CYS-RICH.

FT DOMAIN 720 987 PROTEIN KINASE.

FT NP_BIND 726 734 ATP (BY SIMILARITY).

FT BINDING 753 753 ATP (BY SIMILARITY).

FT ACT_SITE 845 845 BY SIMILARITY.

FT DISULFID 195 204 BY SIMILARITY.

FT DISULFID 199 212 BY SIMILARITY.

FT DISULFID 236 244 BY SIMILARITY.

FT DISULFID 240 252 BY SIMILARITY.

FT DISULFID 255 264 BY SIMILARITY.

FT DISULFID 268 295 BY SIMILARITY.

FT DISULFID 299 311 BY SIMILARITY.

FT DISULFID 315 331 BY SIMILARITY.

FT DISULFID 334 338 BY SIMILARITY.

FT DISULFID 511 520 BY SIMILARITY.

FT DISULFID 515 528 BY SIMILARITY.

FT DISULFID 531 540 BY SIMILARITY.

FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.

FT DISULFID 567 584 BY SIMILARITY.

FT DISULFID 587 596 BY SIMILARITY.

FT DISULFID 600 623 BY SIMILARITY.

FT DISULFID 626 634 BY SIMILARITY.

FT DISULFID 630 642 BY SIMILARITY.

FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 658 658 V -> E (IN ONCOGENIC NEU).

FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).

CC SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 86.4%; Score 5914.5; DB 1; Length 1254;
Best Local Similarity 86.4%; Pred. No. 1.1e-307;
Matches 1092; Conservative 59; Mismatches 100; Indels 13; Gaps 4;

QY 1 MELAALCRWGLLLALLPQCAASTOVCTGDMKLRLPASPETHLDMRLHLYQGCVQGNL 60
DB 1 MELAAWCGGLLLALLSPGASGTQVCTGDMKLRLPASPETHLDIVRHLYQGCVQGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQGVLYIAHNOVRQVPLQRLRIVRGTLQFEDNYALVDNG 120
DB 61 ELTYLPANATLSFLQDIOEVQGVLYIAHNOVRQVPLQRLRIVRGTLQFEDNYALVDNR 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSITELKGGVLIQRPOLCYQDITLWKDF---NNFTV 177
DB 121 DPLDNVTATGRTPEGLRELQRLSITELKGGVLIQRPOLCYQDITLWKDFVRKNN-- 177
QY 178 SFWLAVPKVSAHLENRSRACHPCSPMKGSCWGESSEDCQSLRTVCAGGCARCKGPL 237
DB 178 ---QLAPVDID--TNRSRACPPCAPACKDNHCWASPEDCQTLTGTIAPRAVPAARL 231
QY 238 PTDCCHCCCAAGCTGPKHSDDLCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTF 297
DB 232 PTDCCHCCCAAGCTGPKHSDDLCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTF 291
QY 298 GASCVTACPNYSLTDVGSCTLVCPLNHQEVTAEDGTQRCCKSKPCARVCYGLGMEHLR 357
DB 292 GASCVTTCPNYSLTEVGSCTLVCPLNHQEVTAEDGTQRCCKSKSCARVCYGLGMEHLR 351
QY 358 EVRAVTSANIOEFAGCKKIFGSLAFIPSPGDDPASNTAPLOPEQLOVFETLEEITGYLY 417
DB 352 GARAITSANIOEFAGCKKIFGSLAFIPSPGDDPASNTAPLOPEQLOVFETLEEITGYLY 411
QY 418 ISAWPDSLPLDSVFQNLQVIRGRIILHNGAYSILTLQGLGISWLGRLSLRGLSGALIHNN 477
DB 412 ISAWPDSLHDLVSFQNLQVIRGRIILHNGAYSILTLQGLGISWLGRLSLRGLSGALIHNN 471
QY 478 THLCFVHTVPMDQLFRNPQALLHTANRPEDECVGEGLAGHOLCARGHCWGPQTCVNC 537
DB 472 THLCFVHTVPMDQLFRNPQALLHTANRPEDECVGEGLAGHOLCARGHCWGPQTCVNC 531
QY 538 SQFLRGQECVBECLVQLGIPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVCAHYK 597
DB 532 SHFLRGQECVBECLVQLGIPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVCAHYK 591
QY 598 DPPFCVAPCPGKPDLSVMPITWKPDEEGACQPCINCTHSCVDLDDKGCAPAEQASPL 657
DB 592 DSPFCVAPCPGKPDLSVMPITWKPDEEGACQPCINCTHSCVDLDDKGCAPAEQASPA 651
QY 658 TSIVSAVVGILLVVLGVVFGILIKRRQKIRKYMRLLOETELVEPLTPSGAMPNQAQ 717
DB 652 TSIIATVVGILLVVLGVVFGILIKRRQKIRKYMRLLOETELVEPLTPSGAMPNQAQ 711
QY 718 MRILKETELRKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLRNTPSKANKELDEA 777
DB 712 MRILKETELRKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLRNTPSKANKELDEA 771
QY 778 YVMAGVSPYVSRLLIGICLTSTVQLVLTQMPYVGLLDHVRNRRGLSGODLLNWCQIAK 837
DB 772 YVMAGVSPYVSRLLIGICLTSTVQLVLTQMPYVGLLDHVRNRRGLSGODLLNWCQIAK 831
QY 838 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETBYHADGGKVPKMMAL 897
DB 832 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETBYHADGGKVPKMMAL 891
QY 898 ESILRRRTHOSDWSYGVYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPTIDV 957
DB 892 ESILRRRTHOSDWSYGVYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPTIDV 951
QY 958 YMIWVKCWMIDSECRPRELVSEFSRMARPOQFVVIQNEIDLPASPLDSTFYFSLLED 1017
DB 952 YMIWVKCWMIDSECRPRELVSEFSRMARPOQFVVIQNEIDLPASPLDSTFYFSLLED 1011

OY 1018 DMGDLVDAEYLVPOGFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEAR 1077
 DB 1012 DMGDLVDAEYLVPOGFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEAR 1071
 OY 1078 RSLAPSEGAGSDVFDGLGMAAGKQLSLTHDPSPLQRYSEDPTVPLPSETDGYVAPL 1137
 DB 1072 RSLAPSEGAGSDVFDGELGMAATKGPQISPRDLSPQLQRYSEDPTVPLPSETDGYVAPL 1131
 OY 1138 TCSPOEYVNOVDVPOPSREGPLPAARPAATLERAKTILSPGKGVVUKDVFAGGAV 1197
 DB 1132 ACSPOEYVNOQEVRRPQPLTPEGPLPVRPAATLERAKTILSPGKGVVUKDVFAGGAV 1191
 OY 1198 ENPEYLTPOGGAPOPHPPAPSPAFDNLNLYWDQDPPERCAPPSTFKGTPTAENPEYLG 1257
 DB 1192 ENPEYLVPRGSSASQPH-PPALCPAFDNLNLYWDQDPSERGPSNPTEGPTAENPEYLG 1250
 OY 1258 DVPV 1261
 DB 1251 DVPV 1254

RESULT 4
 EGFR_HUMAN STANDARD; PRT; 1210 AA.
 ID EGFR_HUMAN STANDARD; PRT; 1210 AA.
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMD9; Q9UMG5; Q92795; Q00732;
 AC O00688; Q9B2S2; Q9H2C9; Q9G2X1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Liberman T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta.";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Maihle N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Garti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer.";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RC TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramanian S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Maihle N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.";
 RL Genomics 71:1-20(2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
 RA Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,
 RA Maihle N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells.";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells.";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells.";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription.";
 RL Oncogene Res. 1:375-396(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis.";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor


```

QY 959 MIMVKCMIDSECRFRELVSFMRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLED 1017
DB 945 MIMVKCMIDADSRKFKRELIIIEFSKMRDPQRYLVIOGDERMHLPSPTDSFYRALMDE 1004
QY 1018 DMGDLVDAEYLYVPOQGFCCDDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEERAP 1077
DB 1005 EDMDDVDVADEYLYVPOQGF-----SSPSTS 1030
QY 1078 RSLAPSGAGSDVFDGDLGMAAGKLOSLPHDPSPLQRYSEDTPVLPSET--DGVA 1135
DB 1031 RTPLLSSLSATSN--NSTVACIDRNLQSCPKEDSFLOYSDDTGTALTEDSIDDTFL- 1087
QY 1136 PLTCSPOEYVNPQVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVYVFAFGG 1195
DB 1088 -----PVPEYINQ-SVPKRPAGSVQNPVHNPQNP-----APSRDPHYQD--PHST 1131
QY 1196 AVENEPEYL-TPOGGAAPQHPHPPAFSPAFDNLVYNDQ-----DP-----PERGA 1238
DB 1132 AVNPEYLLNTVQ-----PTCVNSTFDSPAHWAQKSGHQSISLDNPDYQDFFPKPEAK 1182
QY 1239 PPSTFKGTPTAENPEYL 1255
DB 1183 PNGIFKGS-TAENAEYL 1198

RESULT 5
ID _EGFR MOUSE STANDARD; PRT; 1210 AA.
AC 001279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
(bek/KGFR) gene.";
RL Oncogene 7:1957-1962 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59 (1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
receptor tyrosine kinase.";
RL Genes Dev. 8:399-413 (1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Iax I., Ullrich A., Schlessinger J., Givol D., Morse B.;

```

```

RT "Comparison of EGF receptor sequences as a guide to study the ligand
binding site.";
RL Oncogene 6:673-676 (1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eisinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the EGF-receptor leads to
dimerization, internalization of the EGF-receptor complex,
induction of the tyrosine kinase activity, stimulation of cell DNA
synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X78987; CAA55587.1; -
DR EMBL; U03425; AAA17899.1; -
DR EMBL; X59698; CAA42219.1; -
DR EMBL; L06864; AAA53029.1; -
DR EMBL; Z12608; CAA78249.1; -
DR HSSP; P11362; 1FGK
MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
FT SIGNAL 1 24
FT CHAIN 25 1210
FT DOMAIN 25 647
FT TRANSMEM 648 670
FT DOMAIN 671 1210
FT REPEAT 75 300
FT REPEAT 390 600
FT DOMAIN 1028 1071
FT DOMAIN 714 981
FT NP_BIND 720 728
FT BINDING 747 747
FT ACT_SITE 839 839
FT DISULFID 190 199
FT DISULFID 194 207
FT DISULFID 215 223
FT DISULFID 219 231
FT DISULFID 232 240
FT DISULFID 236 248
FT DISULFID 251 260
FT DISULFID 264 291
FT DISULFID 295 307
FT DISULFID 311 326
FT DISULFID 329 333

```


| FT | NP_BIND | 724 | 732 | ATP (BY SIMILARITY). |
|---|----------|---|------------|--|
| FT | BINDING | 751 | 751 | ATP (BY SIMILARITY). |
| FT | ACT_SITE | 843 | 843 | BY SIMILARITY. |
| FT | DISULFID | 189 | 197 | BY SIMILARITY. |
| FT | DISULFID | 193 | 205 | BY SIMILARITY. |
| FT | DISULFID | 213 | 221 | BY SIMILARITY. |
| FT | DISULFID | 217 | 229 | BY SIMILARITY. |
| FT | DISULFID | 230 | 238 | BY SIMILARITY. |
| FT | DISULFID | 234 | 246 | BY SIMILARITY. |
| FT | DISULFID | 249 | 258 | BY SIMILARITY. |
| FT | DISULFID | 262 | 289 | BY SIMILARITY. |
| FT | DISULFID | 293 | 304 | BY SIMILARITY. |
| FT | DISULFID | 308 | 323 | BY SIMILARITY. |
| FT | DISULFID | 326 | 330 | BY SIMILARITY. |
| FT | DISULFID | 503 | 512 | BY SIMILARITY. |
| FT | DISULFID | 507 | 520 | BY SIMILARITY. |
| FT | DISULFID | 523 | 532 | BY SIMILARITY. |
| FT | DISULFID | 536 | 552 | BY SIMILARITY. |
| FT | DISULFID | 555 | 569 | BY SIMILARITY. |
| FT | DISULFID | 559 | 577 | BY SIMILARITY. |
| FT | DISULFID | 580 | 589 | BY SIMILARITY. |
| FT | DISULFID | 593 | 614 | BY SIMILARITY. |
| FT | DISULFID | 617 | 625 | BY SIMILARITY. |
| FT | DISULFID | 621 | 633 | BY SIMILARITY. |
| FT | MOD_RES | 1162 | 1162 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). |
| FT | MOD_RES | 1188 | 1188 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). |
| FT | MOD_RES | 1258 | 1258 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). |
| FT | MOD_RES | 1284 | 1284 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). |
| FT | CARBOHYD | 138 | 138 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 174 | 174 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 181 | 181 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 253 | 253 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 358 | 358 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 410 | 410 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 473 | 473 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 495 | 495 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 548 | 548 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 576 | 576 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 620 | 620 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | VARSPLIC | 626 | 648 | NGPTSHDCIYPTWGHSTLPHQA -> IGSSEDICGLMD (IN ISOFORM JM-B). |
| SEQ | SEQUENCE | 1308 AA; | 146807 MW; | 584AE80985D88761 CRC64; |
| Query Match 43.4%; Score 2972.5; DB 1; Length 1308; | | | | |
| Best Local Similarity 45.1%; Pred. No. 5.3e-151; | | | | |
| Matches 613; Conservative 185; Mismatches 369; Indels 191; Gaps 31; | | | | |
| Qy | 9 | WGLLLALLPGAA-----STOVCTGTDKMLRASPETHLMDLRHLRYGCGQVQGNLBTY | 64 | |
| Db | 8 | WVWVSLVAAGTVQPSDSQSCVCAGTENKLSLSLDLEQQYRALRKYKYENCEVWVGNLEITS | 67 | |
| Qy | 65 | LPTNASLSFLQDIQEVGYVLIAHNOVROVPLQRLRIVRGTOGLFEDNYALAVLDNGDPLN | 124 | |
| Db | 68 | IEHNRDLSFLRSVREVTGYVVALNQRYLPLENRLRIIRTKUYEDRYALAFILNYRKDG | 127 | |
| Qy | 125 | NTPTVTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYODTILWKDF-----NNFTVS | 178 | |
| Db | 128 | NF-----GLQELGKNLTEILNGVGVVDQNKFLCYADTIHWQDIVRNWPNSLTL- | 177 | |
| Qy | 179 | FWLRVPKVSASHLENRSRACHPCSPMCKSRCHGESSEDCQSILTRTVCAGC-ARCKGPL | 237 | |
| Db | 178 | -----VST-----NGSSGCGRCHKCTG-RCHGPTENHCOTLTRTVCAQCDCGRCYGPY | 225 | |
| Qy | 238 | PTDCHEQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTVNTDTFESMPNPEGRYTF | 297 | |
| Db | 226 | VSDCCHRECAAGCGSPKDTDCPLAFNDSGACVTCQPTQTFVTNPTTFQLEHFNNAKIT | 285 | |
| Qy | 298 | GASCVTACPNYILSTDVGSCTLYCPLHNQVBTAEADGTQRCCKSKPCARVCYGLGMEHLR | 357 | |
| Db | 286 | GAFCVKCKCPHNFV-VDSSSCVCRACPSCKMEV-EENGIKMKCPCTDICPKACDGI GTGSLM | 343 | |
| Qy | 358 | EVRAVTSANIQEFAGCKKI FGS LAFLPESFDGDPASNTAPLQPEQLQVETLEEITCYLY | 417 | |

```
Db 344 SAQTVDSSNIDKFNCTKINGNLIFLVTGIHGDYPNAIEAIDPEKLVNVRTVREITGFLN 403
Qy 418 ISAWPDSLPDLSVFONLQVIRGRIHLHNGAYSITLQGLGISWGLRSLRGLSGLALIHNN 477
Db 404 IQSWPNPNTDFSVFNLVITIGRVLYSLGLSLILKQOGITSLQFQSLKEISAGNIYITDN 463
Qy 478 THLCFVHTVPMDLFRNPHOALLHTANRDECEVCGELACHOLCARHGCWGPPTQCVNC 537
Db 464 SNLCYHTINWTLTSTINQRIIVIRDNKAENCAETAGWCVNHLCSGDCGCGPPOQCUSC 523
Qy 538 SOFLRQBCVECEVRVLOGLPREYVNAHCLPCHPECQP-QNGSVTCFGEADQCACAHY 596
Db 524 RFRSGRICIESCNLYDEFEFENGISICVECDPQCEKXMEDGLLCHGPGPNDCTKSHF 583
Qy 597 KDPPFCVARCSPGVKPDLSYMPDIWKFPPDEEGACQPCPNCTHSCVDLDDKGC----- 648
Db 584 KDGPNCEKCPDGLQANSF--IFKYADPDRECHPCNCTQCGNGPTSHDCIYVPWTGH 641
Qy 649 ---PAEQRASPLTSTVSAAV-GILLVVLGVVFGILIKRQOKIRKYTMRRLLQTELVE 704
Db 642 STLPOHAR-TPL--IAAGVIGGLFVLIVGLTFVAVVRKSIK-KKRALRRFL-ETELVE 696
Qy 705 PLTPSGAMPNQAMRILKETELKRVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLREN 764
Db 697 PLTPSGTAPNQAQRLKETELKRVKVLGSGAFGTYYKGIWPEGETVKIPVAIKILNET 756
Qy 765 TSPKANKEILDAYMAGVSGPYVSRLLIGLITSTVQLVTQMLPYGCLLDHVRENRRIG 824
Db 757 TGPKANVFMEALIMASMDHPLVRLVGLVCLSPITQLVTQMLPRHGCLEYYVHEKDNIG 816
Qy 825 SQDLLNWCQAKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVH 884
Db 817 SOLLNWCQVAKGMVLEERLVHRDLAARNVLKSPNHVKITDFGLARLLGEGEKEYN 876
Qy 885 ADGKVPKKNMALESILRRRFTHQSDVMSYGVVWELMTFGAKPYDGIPAREIPDLLEKG 944
Db 877 ADGKVPKKNMALECIHVRKFTHQSDVMSYGVVWELMTFGKPYDGIPAREIPDLLEKG 936
Qy 945 ERLPQPPCTIDVYMWKCMIDSECRPRELVSEFSRMARDQRFVWQNEQ-LGPA 1003
Db 937 ERLPQPPCTIDVYMWKCMIDADSRPKFELAAEFMRARDQRYLVLIQGDQRMKLP 996
Qy 1004 SPLDSTFRSLDEDDMDGLVDAEYLVLPQOQFFCPDPAGAGVMVHRHSSSRSGG 1063
Db 997 SPNUSKFFONLDEEDLMDMAEYLVLP-QAFNIPP-----IYTSRARDISNRS--- 1046
Qy 1064 DLTGLPSEBEAPRS-----PLAP-SEGAGSDVFDG 1094
Db 1047 --EIGHSPPPAYTPMSGNQFVYVROGGFAAEGSVVPYRAPTSTIPEAPVAQATAEIDD 1104
Qy 1095 DLGGAAGKLOSLPHTDPSLQRYSEDTVPPLS-----ETDGYVAPLTCSPQPEYVN 1147
Db 1105 SCCNGTLRKVAPHVQEDSSSTORYSADPTVPAPERSPRGELDEGYMTFMRDKPKQEYLN 1164
Qy 1148 QPDRVPQPPSPREGPLPAAPAGATLERAKTILSPKNGVWDVFAFGGAVENPEYLTQ 1207
Db 1165 PV-----ENPFVSR-----KNGDLQ-----ALONPEYHNASN 1193
Qy 1208 GAAPQHPHPPA-----FSPAFDNLVYWDQOP 1233
Db 1194 G-----PPKADEYVNEPLYLNTFANTLGKAEYLNKNLSMPEKAKKAFDNPVWNHSL 1247
Qy 1234 PERGA--PPSTFKGTPT-----AENPEYL 1255
Db 1248 PPRSTLQHPDYLQYESTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 7
ID ERB4 RAT STANDARD; PRT: 1308 AA.
AC Q62956; Q92N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
```

```
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NCBI_TaxID=10116;
OX 111
RN 1
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RN J. Biol. Chem. 273:10261-10269(1998).
RL 2
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN 3
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9010624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC 2. NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF- $\alpha$ , AND AMPHIREGULIN (BY SIMILARITY).
CC -! CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -! SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -! TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -! PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -! SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041838; AAC08899.1; --
CC EMBL; U52531; AAC53051.1; --
CC HSRP; P11362; 1FGK.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP_motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD0000001; Euk_pkinase; 1.
```

DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transphosphatase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 PROTEIN KINASE.
FT DOMAIN 718 985 NP BIND (BY SIMILARITY).
FT NP BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 PHOSPHORYLATION (AUTO-).
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match 43.2%; Score 2959; DB 1; Length 1308;
Best Local Similarity 45.2%; Pred. No. 2.8e-150;
Matches 612; Conservative 191; Mismatches 382; Indels 170; Gaps 30;

QY 1 MELA-ALCRWGLLL--ALLPPGAASQVCTGTDMKRLPASPETHLDMLRLHLYGQCQVQ 57
DB 1 MKLATGLWVWGLSLVAARTVQPSASQSVCAGTENKLSLSLDLEQQYRALRKYKYNCEVVM 60
QY 58 GNLETLTYLTNLSLSFLQIOEQGVYLAHNQVQVPIQLRIVRGTOLFDNYALAVL 117
DB 61 GNLEITSIBHNRDLSFLRSIREVTGYLVVALNQFRLPLENRIIRGTLKLYEDRYALAIF 120
QY 118 DNGDPLNNTPTVTPGSPGLRLQLRLSLTEILKGGVLIQNPOLCYQDTILWKDFNFTV 177
DB 121 LNYRKDNF-----GLQELGLKVLTEILNGGVYVDQNKFLCYADTIHWQDI----- 166

QY 178 SFWLRVPKVSASHLENR--SRACHPCSPMKGRSCRMGESSEDCQSILTRTVCAAGC-ARCK 234
DB 167 ---VRNPWPSNMTLVSTIGSSGGRCHKSGCTG-RCWGPTEHNCQTLTRTVCAEQDGRCY 222
QY 235 GPLPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTVNTDTTFESMPNPEGR 294
DB 223 GPVSDCHRECAAGCGSGPKDTCFACMNFNDSGACVTCQPTQFVFNVTTFQLEHFNPAK 282
QY 295 YTFGASCVTACPVNYLSTDVGSCTLVCPLHNOEVTAEADGTORCEKSKPCARVCYGLGME 354
DB 283 YTYGAPCVKKCPHNFV-VDSSCVCRACPSKMEV--EENGIMKCPKCTDICTPKACDGI 340
QY 355 HLREVRVATSANIOEFAGCKIFGSLAFIPESFGDPASNTAPLQPEQLQVFFLEBITG 414
DB 341 SLMSAQTVDSNIDKFINCTKINGNLJFLVTGHHGDPYNAIDAIPEKLNFTVREITG 400
QY 415 YLYISAWPDSLPDLVSFQNLQVIRILHNGAYSLTLOGLGISWLGRLSRLGSLGLALI 474
DB 401 FLNIQTWPPNMTDFSVFSLNLTIGRVLVSGLSLLILKQOGITSLQFSLKEISAGNIYI 460
QY 475 HNTHLCFVHTVPDOLFRPHQALLHTANRPEDECVGSEGLACHOLCARGHCWGPPTQC 534
DB 461 TNSNLCCYYHTINWTLFTSTVNQIRIVIRDNRAENCTAEGMVCNHLCSNDGCGPDPQC 520
QY 535 VNCQSLRQOEVEECRVLQGLPREYVNAHCLPCHPEQCP-QNGSVTCFGEADQCVAC 593
DB 521 LSCRRFSRGKICIESCNLYDGEFRENGSICVECDSCQCEKMEGGLTCHGPGPDNCTKC 580
QY 594 AHYKOPPFVCVARGCPGKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGC----- 648
DB 581 SHFKDGNPCVCKPDVLOGANSF--IFKYADQDRECHPCPNCTQCGNGTSHDCIYYPW 638
QY 649 -----PAQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMERLLOETE 701
DB 639 TGHSTLPQHAR-TPL--IAAGVIGGLFILLVWALTFAVYVRRKSIK-KKRALRFL-ETE 693
QY 702 LVEPLTPSCAMPNQAMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVL 761
DB 694 LVEPLTPSCAPNQAOQLRILKETELRKVKVLSGAGFTVYKGIWVPEGETVKIPVAIKIL 753
QY 762 RENTSPKANKEILDEAYVMAGVSPVSRLLGLTSTVOLVTQMLMPYGCLLDHHVRENRG 821
DB 754 NETTGPKANVEFMDEALIMASVDHPLVRLGLVCLSPITQLVTQMLPHGCLLEYVHEHKD 813
QY 822 RLGSDDLNNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDET 881
DB 814 NIGSQLLNWCVOIAKGMVLEBERRLVHRDLAARNVLKSPNHVKITDFGLARLLGDEK 873
QY 882 EYHADGGKVPICKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLL 941
DB 874 EYNADGGKVPICKMALECIHYRKFTHQSDVMSYGVTVWELMTFGKPYDGIPTREIPDLL 933
QY 942 EYGERLPQDPICITIDVYMIWVKWIMIDSECRPRFRELVSFESRMARDPQRFVVIQNE-L 1000
DB 934 EYGERLPQDPICITIDVYIYVMVKWIMIDADSRPFKEALAAEFMRMARDPQRYLVIOGDDRM 993
QY 1001 GRASPLDSTFYRSLLEDDEDDMDGLVDAEEYLVPOQGFCCDP----- 1041
DB 994 KLPSNDSKFFQNLDEEDLEDDMDABEYLVLP-QAFNIPPIYTSRTRIDSNRSEIGHSP 1052
QY 1042 ----APGAGMYYHRRHSSTRSGGDLTLGLEPSEEAAPRSPPLASSEGAGSVDFQDGLG 1097
DB 1053 PPAYTPMSGSQFYQDGGFATQOG---MPMPYATTTSTTPEAPVA--QQAETAEFDDSCC 1107
QY 1098 MGAAGKLOSLPHTDPSPLQRYSEDPVPLPS-----ETDGYAPLTCSPQPEYVNPQD 1150
DB 1108 NGTLKRPVPHVQEDSSTORYSADPTVFAPERNPRAELDEEGYMTBMHDKPKQEYLNPE 1167
QY 1151 VEPQPPSPREGPLPAARPAGATLERAKTSLSPKNGVVKVDFVFGGAVENPEYLTPOGGAA 1210
DB 1169 -----ENPFVSR-----KNGDLQ-----ALDNPYHSASSG-- 1194
QY 1211 POPHPPPA-----FSPAFDNLVYWDQDPPER 1236

```

Db 1195 ----PPKADEYVNEPLYNTFTNALGNAYMNSLLSVPEKAKAFDPDITWNSLPPR 1250
QY 1237 GA--PPSTFKGPT-----AENPEYL 1255
Db 1251 STLQHPDYVLEQYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 8
ID XMRK_XIPMA STANDARD; PRT; 1167 AA.
AC P13388;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TU.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI TaxID=8083;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Wittbrodt J., Adam D., Malitschek B., Mauelel W., Raulf F.,
RA Telling A., Robertson S.M., Scharl M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT inducing Tu locus in Xiphophorus.";
RL Nature 341:415-421(1989).
RN [2]
RP REVISION TO 515.
RA Scharl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16891; CAA34770.2; -
DR PIR; S06142; S06142.
DR HSSP; P13362; IFGK.
DR InterPro; IPR000494; EGF_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PR000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 25
FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

```

```

FT DOMAIN 26 642
FT TRANSMEM 643 665
FT DOMAIN 666 1167
FT NP_BIND 710 977
FT BINDING 716 724
FT ACT_SITE 743 743
FT DISULFID 835 835
FT DISULFID 199 212
FT DISULFID 220 228
FT DISULFID 224 236
FT DISULFID 237 245
FT DISULFID 241 253
FT DISULFID 256 265
FT DISULFID 269 296
FT DISULFID 300 311
FT DISULFID 315 330
FT DISULFID 333 337
FT DISULFID 504 513
FT DISULFID 508 521
FT DISULFID 524 533
FT DISULFID 537 553
FT DISULFID 556 569
FT DISULFID 560 577
FT DISULFID 593 615
FT DISULFID 618 626
FT DISULFID 622 634
FT CARBOHYD 114 114
FT CARBOHYD 144 144
FT CARBOHYD 201 201
FT CARBOHYD 356 356
FT CARBOHYD 365 365
FT CARBOHYD 398 398
FT CARBOHYD 417 417
FT CARBOHYD 501 501
FT CARBOHYD 576 576
FT CARBOHYD 621 621
SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 39.3%; Score 2690.5; DB 1; Length 1167;
Best Local Similarity 45.1%; Pred. No. 4.8e-136;
Matches 576; Conservative 165; Mismatches 385; Indels 151; Gaps 30;

QY 4 AALCRGWLALLPQGAAS-----QVCTGDMKRLPASPETILDMRLHYQGCQVQGN 59
Db 8 AALQ--LLVLISRCSTDPDRKVCQGTSNQMTM---LDNHYLKKKHYGSCNVLEN 62
QY 60 LELYPTNASLSFLQDIOEVGYVLIAHNVQVPLQRLRIVRGTLQFEDNYALAVLDN 119
Db 63 LEITYTQENQDLSFLQIOEVGYVLIAMNEVSTIPLVNLRLIRGNLYEGNFTLLVMSN 122
QY 120 GDPLNNTTPTGASPGGLRELRLSLTEILKGGVLIQRNPOLCYODTILWKDF-----NNF 175
Db 123 YQK-PPSSP--DVYQVGLKQLQSLNLTLSGGVKSHPNLLCNVETIWMMDIVDKTSNP 179
QY 176 TVSFWRVPKVSASHLENRSRACHPCSPMKSGRCWGESSEDCQSILTRVCAGGC-ARCK 234
Db 180 TML--IPHAF-----ERQCKDHGCVNGSWAPGPGHCQKFTKLLCAEQCNRRCR 229
QY 235 GPLPTDCHEQCAAGCTGPKHSKSLCLAHFNHSGICELHCPALVTYNTDTFSPMPNPEGR 294
Db 230 GPKPIDCCNEHCAGCTGPRATDCIACRDFNDGCTKDTCPPPKIYDIVSHQVWDNPNIK 289
QY 295 YTFGASCVTACPNYVNSTDVGSCITLVCPLHNQEVTAEDGTQRCCKSPCARVCYGLGNE 354
Db 290 YTFGAACVKCEPNSYVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKCDGIGIG 347
QY 355 HLREVRVAVTSANIOEFAGCKIFGSLAPLPSFDGDPASNTAPLOEQLQVFPETLEEITG 414
Db 348 SLSNTIAVNSNIRSFNCTKINGDIIILNRNSFEGDPHYKGTMDPEHLMLTWTVEITG 407
QY 415 YLYTSAMPDLSPLSVFQNLQVIRILHNGAYS-LTLQGLIGISWLGRLSRLSRELGSGLAL 473

```



```
Db 408 YLVIMWMPENMTSLSFQNLRIIRTTFSRGFSFVVQVRHLQWLGLSLKEVSAGNVI 467
Qy 474 IHHNTHLCFVHTVPDOLFRNPHOALLHTANRPEDECVGEGLACHOLCARGHCWGPGPTQ 533
Db 468 LKNTQLQRLVANTINRRRLFRSEDSLEYDART-----ENQTCNNECEDGCGWGPPTM 520
Qy 534 CYNCSOFLRGQCEBECRVLQGLPREYVNRHCLPCHPECCQPNQSVTCFGPEADQCVAC 593
Db 521 CVSCSLHVDGRGRCVASCNLLQGEPREAQVDGRCVQCHQECVLQTDLSLTCYGGPANCSSK 580
Qy 594 AHYKDPFVCVAPSGVKPDLSPYPIWKPFDEGACQPCPINCSTHSCVDLDDKGCAPR 653
Db 581 AHFDQGPQICPRPHGILGDGTL-IWKYADKRGQCQCHQNCQTCGCGPGLSGCRGD-I 638
Qy 654 ASPLTSIVSAVGVILLVVGVVFGILIKRRROOKIRKYTMRLQLQELVEPLTPSGAMP 713
Db 639 VSHSSLAUGLVGLLITVALLIVLLRRRIK-RKRTIRCLLQEKELVEPLTPSGQAP 697
Qy 714 NOAQRILKETELRKVKVLGSGAFGVYKGIWIPGENVKI PVAIKVLRENTSPKANKEI 773
Db 698 NOAFLRIKETEFKDRVLGSGAFGVYKGLWNPGENIRIPVAIKVLEATSPKVNQEV 757
Qy 774 LDEAYVACGSPYSRLIGLICLTSTVOLVTQLMPYGCILLDVRNHRGLGSDLLNWC 833
Db 758 LDEAYVMSVDHPHVCRLGICLTSAVOLVTQLMPYGCILLDVRNHRGLGSDLLNWC 817
Qy 834 QIAGKMSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHYHADGKVPK 893
Db 818 QIAGKMYLEERHLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHYHADGKVPK 877
Qy 894 WMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPIC 953
Db 878 WMALESILQWTVTHQSDVMSYGVTVWELMTFGSKPYDGPAREIASVLENGERLPQPPIC 937
Qy 954 TIDVYIMVKWIMIDSECPRELIVSESRWARDPQRFVIONEDLGPASPLDSTFYRS 1013
Db 938 TIEVYMIILKWMIDSPRELIVSESRWARDPQRFVIONEDLGPASPLDSTFYRS 994
Qy 1014 LLEDDDMGLVDAEYLVPOQGFCCPDAPGAGGMVHRHRSSTSGGGDLTLGLPESE 1073
Db 995 LLSSDD--DVDADEVLLPYKRI-----NQGS----- 1020
Qy 1074 EAPRSLPAPSGAGSDVFDGLMGAAKGLSLPHDPSLPQRYSEDTV-PLPSETGD 1132
Db 1021 -----EPCIPPTGH-----PVRENSITLRNISDPTQNALEKDLGD 1055
Qy 1133 YVAPLTCSPQEVNQPVRPQP-----PSPRE-----GLP-AARAGATLERAKTL 1179
Db 1056 H-----EYVNPQGETSSRLSDIYNPNYEDLTDGMPVSLSSQBAETNFSRPEYL 1105
Qy 1180 SPKGKGVKDVFAFGGAVENPEYLTPOGGAAPQHPHPAPFAPFNLYYWDQDPPRGPAP 1239
Db 1106 NTNQNSL---PLVSSGMDPDY---QAG-----YQAF-----LPQTGAL 1140
Qy 1240 PSTFKGTPTAENPEYLG 1256
Db 1141 TGNMGFLPAENLEVLG 1157
```

RESULT 9

```
ERB3_HUMAN
ID ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kraus M.H., Ising W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA Plowman G.D., Shitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Whitney M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RA TISSUE=Placenta;
RC MEDLINE=93282922; PubMed=7685162;
RA Kato M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M29366; AAA35790.1; -
CC EMBL; M34309; AAA35979.1; -
CC EMBL; S61953; AAB26935.1; -
CC PIR; A36223; A36223.
CC HSSP; P11362; 1FGK.
CC Genew; HGNC:3431; ERBB3.
CC MIM; 190151; -
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC ProDom; PD0000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKC; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
```

KW Alternative splicing, 19
 FT SIGNAL 20 1342
 FT CHAIN 20 643
 FT DOMAIN 20 643
 FT TRANSMEM 644 664
 FT DOMAIN 665 1342
 FT DOMAIN 709 966
 FT NP BIND 715 723
 FT BINDING 742 742
 FT ACT_SITE 834 834
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235
 FT DISULFID 231 243
 FT DISULFID 246 255
 FT DISULFID 259 286
 FT DISULFID 290 301
 FT DISULFID 305 320
 FT DISULFID 323 327
 FT DISULFID 500 509
 FT DISULFID 504 517
 FT DISULFID 520 529
 FT DISULFID 533 549
 FT DISULFID 552 565
 FT DISULFID 556 573
 FT DISULFID 576 585
 FT DISULFID 589 610
 FT DISULFID 613 621
 FT DISULFID 617 629
 FT CARBOHYD 126 126
 FT CARBOHYD 250 250
 FT CARBOHYD 353 353
 FT CARBOHYD 408 408
 FT CARBOHYD 414 414
 FT CARBOHYD 437 437
 FT CARBOHYD 469 469
 FT CARBOHYD 522 522
 FT CARBOHYD 566 566
 FT CARBOHYD 616 616
 FT VARSPPLIC 141 183
 FT VARSPPLIC 184 1342
 FT CONFLICT 560 560
 FT CONFLICT 1064 1064
 SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;

Query Match 35.4%; Score 2427.5; DB 1; Length 1342;
 Best Local Similarity 40.5%; Pred. No. 5.7e-122;
 Matches 533; Conservative 192; Mismatches 457; Indels 135; Gaps 32;

QY 10 GLLALLPPGAA--STQVCTGTDMLRLPASPTHLDMLRLHYQGVVQGNLELTLYLPT 67
 DB 11 GLLFSLARGSEVGNQAVCPFTGLSLVTGDAENQVTLKLYERCEVVMGNLEILVTCH 70
 QY 68 NASLUSFLODIOEVOQYVLIANNQVQLRQIRIVRGTLQFEDNYALAVLDNGDPLNNTT 127
 DB 71 NADLSFLOWIREVTGYVLVANNEFTSLPLNLRVVRGVQVYDGKFAIFW-----LNYNT 125
 QY 128 PVTGASPGGLRELQRLSTEILKGGVLTORNPOLCYQDTILWKDFNNFTVFWLRVPKVS 187
 DB 126 ----NSSHALKQLRUTQTEILSGGVYTEKNDKLCMDTIDWRDVR-----DRDA 172
 QY 188 ASHLENRRACHPSCPMCKGRGCMGSESDCOSLRTVTCAGGC-ARCKGPLPTDCCHQC 246
 DB 173 EIVVKDNGRSCPPCHEVCKG-RWCGPSEDCOTLTKTICAPQCNGHCFGPNQCHDEC 231
 QY 247 AAGCTGPKHSCDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 306
 DB 232 AGGCGSGPQDTCFACRHFNDGACVPRCFQPLVKNLTFQLEPNPHHTKYQYGGVCVASCP 291

QY 307 YNYLSTDVSGCTLVCPHLHQEVTAEDGTORCEKSKPCARVCVGLGMEHUREVRAVTSAN 366
 DB 292 HNFV-VDTQTSVCVRACPPDKMEVD-KNGLKMXCEPCGGLCPKACBGTGSG--SRFTQTDSSN 347
 QY 367 IQEFAGCKKI FGSFLAPLPESEFDGDPASNTAPLOPEQLOVFEETLEEITGYLYISAWPDSL 426
 DB 348 IDGFVNCTKILNLDLFLITGLNGDPWHKIPALDPEKLNFRVREITGYLNIOSWPPHMH 407
 QY 427 DLSVFQNLQVIRGRIHNGAYS-LTLQGLGISWLGSLRSLRELGSLALIHNNTHLCFVHT 485
 DB 408 NFSVFSNLTTIGRSLYNRGFSLLIMKNLNVTSLSGFRSLKEISAGRIYISANQOLCVHHS 467
 QY 486 VPDQLFRNPQOALLHTA-NRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRQ 544
 DB 468 LNWTKVLRGPTTEERLDIKHNRPRRDCAEGKVCVCLPCSSGCGWGPQGLSCRNYSRG 527
 QY 545 ECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVA 604
 DB 528 VCVTHCNFLNGEPREFAEAECSCHPECPQMEGTATNGSGSDTCAQCAHFRDGHCHVS 587
 QY 605 RCPGKVPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPAEORA-----SPLTSI 660
 DB 588 SCPHGVLG--AKGPIKYKYPDVQNECRPCHENCTQCGKPELQDCLQTLVIGKTHLTMA 645
 QY 661 VSAVVGILLVVVGVVFGILIKRQOKIR-KYTMRLLOQETELVEPLTTPSGAMPNQAQMR 719
 DB 646 LTVIAG--LWIFMMLGGTFLYWRGRIQNKAMRRYLERGESIEPLDPS-ERANKVYLR 702
 QY 720 ILKETELRKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPRANKILDEAYV 779
 DB 703 IFKETELRKVLGSGVGTGVHKGWVPEGESIKIPVCIKVIDKSGRGSFOAVTDHMLA 762
 QY 780 MAGVSPVYSRLIGLCTSTVOLVTQLMFYGCLLDHVRENRRGLSGQDLNLMCMQIAKGM 839
 DB 763 IGSLDHAHIVRLGLCPGSSQLVQYPLGSLLDHVROHRCALGALPQLLLNMGVQIAKGM 822
 QY 840 SYLEDVRLVHRDLAARNVLVKS PNHVKITDPLGARLLDIDETEHYHADGKVPKIMWALES 899
 DB 823 YYLEEHGMVHRNLAARNVLLKSPSQVADFGVADLLPPDDKQLLYSEAKTPIKMALES 882
 QY 900 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLEKGERLPOPPCTIDVYM 959
 DB 893 IHFGYITHQSDVMSYGVTVWELMTFGABPYAGLRLAEVDPDLEKGERLAOPQICTIDVYM 942
 QY 960 IMVKCMIDSECRPRFELVSEFSRMARDPQRFVVIQNEIDLGA-----SPLDSTFYRSLLE 1016
 DB 943 VMVKCMIDENIRPTFKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEHCLTHKKLE 1001
 QY 1017 DDMGDVLDAEYLVPOQGFPCDPAPGAGGMVHHRHRSSTRSGGDLTLGLEP-SEEE 1075
 DB 1002 EVELEPELDLDLDEAED-----NLATTTLSALSPLVGTLNRRP 1042
 QY 1076 APRSLAPSEAGSDVFDGDLGMAAKGLQSLPTH-DPSLPORYSEDPTVPLP-----S 1128
 DB 1043 GSQSLSPSSGY-MPMNQNLGSCQESAVSGSSERCPVSLH-----PMPRGCLASE 1095
 QY 1129 ETDGYVA-----PLTCSQPOPE-----YVNPQDVRPOPSPREP----- 1162
 DB 1096 SSEGHTVSEAELOEKVSMCRSRSRSPRPGDSAYHSQRSHLLTPVTLPLSPGLEEED 1155
 QY 1163 -----LPAARPAGATLERAKTLSP-GKNGV-----KDVFAFGGAVENPEYLTPOGGAA 1210
 DB 1156 VNGYVMPDTHLKGTPSSREGTLSSVGLSVLGTDEEDED-----EVEYMNRRRRHS 1207
 QY 1211 POPHPPPAFSFAFDNLYYWD-----ODPPERCAPPSTFKGTPTAENPEYL 1255
 DB 1208 P-PHPHPPSSLEELGYEYMDVGVSDLSASLGSTQSCPLHPVIMPITAGTTPDEDYEYM 1263

RESULT 10
 ERB3 RAT
 ID ERB3 RAT STANDARD; PRT; 1339 AA.
 AC Q62759; Q62955;

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERBB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
RT recombinant protein";
RL Gene 165:279-284(1995).
RN (2)
RN REVISIONS TO 85; 513 AND 565.
RP Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuroligins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration";
RL J. Neurosci. 17:1642-1659(1997).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTKA.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U29339; AAC28498.2; -;
CC EMBL; U52530; AAC53050.1; -;
CC HSSP; P11362; 1FGK.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19 POTENTIAL
FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).

| | | | | |
|----|----------|------|------------|-----------------------------------|
| FT | TRANSMEM | 644 | 662 | POTENTIAL. |
| FT | DOMAIN | 663 | 1339 | CYTOPLASMIC (POTENTIAL). |
| FT | DOMAIN | 259 | | CYS-RICH. |
| FT | DOMAIN | 707 | 964 | PROTEIN KINASE. |
| FT | NP_BIND | 713 | 721 | ATP (BY SIMILARITY). |
| FT | BINDING | 740 | 740 | ATP (BY SIMILARITY). |
| FT | ACT_SITE | 832 | 832 | BY SIMILARITY. |
| FT | DISULFID | 186 | 194 | BY SIMILARITY. |
| FT | DISULFID | 190 | 202 | BY SIMILARITY. |
| FT | DISULFID | 210 | 218 | BY SIMILARITY. |
| FT | DISULFID | 214 | 226 | BY SIMILARITY. |
| FT | DISULFID | 227 | 235 | BY SIMILARITY. |
| FT | DISULFID | 231 | 243 | BY SIMILARITY. |
| FT | DISULFID | 246 | 255 | BY SIMILARITY. |
| FT | DISULFID | 259 | 286 | BY SIMILARITY. |
| FT | DISULFID | 290 | 301 | BY SIMILARITY. |
| FT | DISULFID | 305 | 320 | BY SIMILARITY. |
| FT | DISULFID | 323 | 327 | BY SIMILARITY. |
| FT | DISULFID | 500 | 509 | BY SIMILARITY. |
| FT | DISULFID | 504 | 517 | BY SIMILARITY. |
| FT | DISULFID | 520 | 529 | BY SIMILARITY. |
| FT | DISULFID | 533 | 549 | BY SIMILARITY. |
| FT | DISULFID | 556 | 573 | BY SIMILARITY. |
| FT | DISULFID | 576 | 585 | BY SIMILARITY. |
| FT | DISULFID | 589 | 610 | BY SIMILARITY. |
| FT | DISULFID | 613 | 621 | BY SIMILARITY. |
| FT | DISULFID | 617 | 629 | BY SIMILARITY. |
| FT | CARBOHYD | 126 | 126 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 250 | 250 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 353 | 353 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 408 | 408 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 414 | 414 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 437 | 437 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 469 | 469 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 522 | 522 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 566 | 566 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 616 | 616 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CONFLICT | 1028 | 1028 | L -> P (IN REF. 3). |
| SQ | SEQUENCE | 1339 | AA; 147545 | MM; 0AA5F2402BBDF1E CRC64; |

Query Match 34.5%; Score 2364.5; DB 1; Length 1339;
Best Local Similarity 40.7%; Pred. No. 13e-118;
Matches 526; Conservative 174; Mismatches 426; Indels 165; Gaps 35;

| | | | | |
|----|-----|----------------------------|--|-----|
| QY | 3 | LAALCRWGLLLALLPPGAA--- | STOVCTGTDMLRLPASPEHLDMLRLHLYQGCVOVQGN | 59 |
| DB | 7 | LQVLC---- | FLISLARGSEMGNSQAVCPGTGLNGLSVTGDADNQYQTLKYKECEVWNGN | 62 |
| QY | 60 | LELTYLPTNASLFLQDIQEVQGYVLI | AHNQVQVPLQRLRIVRGTQQLFEDNYALVDN | 119 |
| DB | 63 | LEITVLTGHNADLSFLQWIREVTGY | VLVAMNEFSVLPNLRVVRGTQVYDGKFAIVM-- | 120 |
| QY | 120 | GDP LNNTTPTVGASPGCLRELQRL | SLTEILKGGVLIQRPOLCYQDTILWKDFNFTVSF | 179 |
| DB | 121 | ---LNYNT--- | NSSHALRQLKFTQTLTEILSGGVYIEKNDKLCMDITDRD----- | 165 |
| QY | 180 | WLRVPKVSASHL-- | ENRSRACHPCSPMGKSGRCWGESSEDCQSLTRTVACAGC-ARCKGP | 236 |
| DB | 166 | ---IVRVGAIEIVVKNNGANGC | PPCHEVCKG-RCWGGPDDCQLTKTTCAPQCNRCFCGP | 221 |
| QY | 237 | LPTDCCHEQCAAGCTGPKHSDCL | ACLHFNHSGICELHCPALVTYNTDTTFESMPNPEGRYT | 296 |
| DB | 222 | NPQCCHDECAGCGSPQDTCFAC | RFRNDSGACVPCRCPELVYKLTFLQLEPNPHTKYQ | 281 |
| QY | 297 | FGASCVTACPYNYLSTVDSCT | LVCPHMQEVTAEDGTORCEKCKPFCARVCYGLGMEHL | 356 |
| DB | 282 | YGGCVASCAPHNFV-VDQTF | FCVRACPPDKMEVD-KHGLKMCPCGGGLCPKACBGTSGS-- | 337 |
| QY | 357 | REVRVTSANIQEFACKIFGSLA | FLPESFGDGRASNTAPLOEQLOVETLEEITGYL | 416 |
| DB | 338 | SRYQTVDSSNDIGFVNCTKIL | GNLDFLITGLNVDPMHKIPALDPEKLVNFTVREITGYL | 397 |
| QY | 417 | YISAWPDSLPLDSVFNQNLQ | VIQRGRIILHNGAYS-LTLQGLGISWLGRLSRLSGLALIH | 475 |

Db 398 NIOSPPPHMHNFVSFNLTIGRSLYNRGFSLLIMKLNVTSLGFRSLKEISARVYIS 457
QY 476 HNTHLCFVHTVWDQLFRNPQALLHTA-NREDEVCVGEGLACHOLCAHGCWGPPTQC 534
Db 458 ANOQLCYHSLNWTLLRGSPSEERLDIKYDRPLGELCLAEAGKVCDCPLCSGGCGWGPQGQC 517
QY 535 VNCOSQFLGOECVEECRVLOGLPREVVARHCLPCHPECPQNGSVTCFPGPADOCVACA 594
Db 518 LSCRNYSGEVCTHCFNLQGGPREFVHQAQCFSCHEPCLPMEGTSTCNGSGSDACARCA 577
QY 595 HYKDPFFCVARCPGSKVDLSYMPIWKFPDEGACQPCPINCTHSC--VDLDDKGCAPAO 652
Db 578 HFDGPHCVNSCPHGLG--AKGPYKYPDQAQNECPCHENTCQCGPELQDCLGQAEV 635
QY 653 RASPLTISVAVGVLLVVVLGVVGLIKRQKIR-KYTWRRLLQETELVEPIPTPSGA 711
Db 636 LMSKPHLVIAVTG--LAVILMILGGSFLYMRGRIRKRAMRYLGERGSEIEPLDPS-E 692
QY 712 MPNOAQRILKETELRKVKVLGSGAPGVYKGIWIPDGENVKIPVAIKVLRNRETSFKANK 771
Db 693 KANKVLARIFRETELURKULVLSGVGTGVHKGWIPEGESIKIPVCIKVIEDKSGRQSFQ 752
QY 772 EILDYAVMAGVSGPYVSRLLGICLTSTVQLVTQMPYGCLLDHRNENRGLGSDLLNW 831
Db 753 AVTDHMLAVGSLDHAHIVRLGLCPGSSQLVTVPLGSLLDHVHQRHETLGPQLLW 812
QY 832 CWOIAKGSYLEDVRLVRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKVP 891
Db 813 GVOIAKGMVYLEEHSVMVRDLARNVLMKSPSQVQVADFGVADLLPPDDKQLLSEAKTP 872
QY 892 IKWMALESILRRRFTHQSDVNSYGVTVMLMTFGAKPYDGPAREIPDLLEKGERLPQP 951
Db 873 IKWMALESIHFGKYTHQSDVNSYGVTVMLMTFGAEPYAGRLAIPDLLEKGERLAQPO 932
QY 952 ICTIDVYIMVKCMIDSECRPRFRELSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFY 1011
Db 933 ICTIDVYIMVKCMIDENIRTFKELANEFTRMARDPRLYLVIKEAS-GPCTP--PAEE 989
QY 1012 RSLLEDDMGDLVDAEYLVPOQGFCDPDPAGAGMVHRRHSSTRSGGDLTLGLEP 1071
Db 990 PSLVTTKEL-----QEAELEPEL-----DLDLDEA 1015
QY 1072 SEE-----EAPRSLAPSEG-----AGSDVFDGDLGWAAGK 1103
Db 1016 EEEGLATSLGALSILPTTLTRPGSQSLSPSGYMPMNOSSLGEACLDASVAGREQF 1075
QY 1104 LQSLPHTDPSLQRYSEDTVPLPSETDGYV-----APL-----SPOPE--- 1144
Db 1076 SRPISLH-PIPRGR-----PASESSEGHVGTGSAELOEKVSVCRSRSRSPRGRDS 1127
QY 1145 -YVNPQDVRPOPSPREGP-----LPAARPAGATILERAKTLS-P-GKNGYV---- 1187
Db 1128 AYHSORHSLTPTVPLSPGLEEDGNGYVMPDTHLRGASSSREGTLSSVGLSSVLGTEE 1187
QY 1188 --KDVFAFGGAVENPEYLTPQGGAPQPHPP 1216
Db 1188 EDED-----EYEYMNRRKRGSP-PRPP 1209

RESULT 11

EGFR_DROME STANDARD; PRT: 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DN (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GS EGFR OR TOP OR C-ERBB OR DER OR CCI0079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
MEDLINE=94350209; PubMed=8070664;
Clifford R., Schubach T.;
"Molecular analysis of the Drosophila EGF receptor homolog reveals that several genetically defined classes of alleles cluster in subdomains of the receptor protein.";
Genetics 137:531-550(1994).
[2]
REVIEWS.
Clifford R., Schubach T.;
Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
MEDLINE=85124611; PubMed=2982499;
Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
"The Drosophila EGF receptor gene homolog: conservation of both hormone binding and kinase domains.";
Cell 40:599-607(1985).
[4]
SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
STRAIN=Oregon-R; Tissue=Embryo;
MEDLINE=87002474; PubMed=3093080;
Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
"Alternative 5' exons and tissue-specific expression of the Drosophila EGF receptor homolog transcripts.";
Cell 46:1091-1101(1986).
[5]
SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
MEDLINE=99102120; PubMed=9882502;
Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
"Several levels of EGF receptor signaling during photoreceptor specification in wild-type Ellipse, and null mutant Drosophila.";
Dev. Biol. 205:129-144(1999).
[6]
SEQUENCE FROM N.A. (ISOFORM TYPE I).
STRAIN=Berkley;
MEDLINE=20196006; PubMed=10731132;
Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Cocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abril J.F., Agbayani A., An H.-J., Andrews-Efannkoch C., Baldwin D., Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jaatl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasseman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RX STRAIN=Daekwanryeong;
RC MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A *Drosophila* genomic sequence with homology to human epidermal
RT growth factor receptor.";
RL Nature 314:178-180(1985).
RN [8]
RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS
RX MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "Interallelic complementation among *DER/fib* alleles: implications for
RT the mechanism of signal transduction by receptor-tyrosine kinases.";
RL Genetics 129:191-201(1991).
RN [9]
RP REVIEW
RX MEDLINE=97248481; PubMed=9094709;
RA Perkinson N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the *Drosophila*
RL EGF receptor.";
CC Cell 89:13-16(1997).
CC -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF052754; AAC08536.1; -;
DR EMBL; AF052753; AAC08536.1; JOINED.
DR EMBL; AF052754; AAC08535.1; -;
DR EMBL; AF052752; AAC08535.1; JOINED.
DR EMBL; K03054; AAA51462.1; -;
DR EMBL; K03417; AAA51460.1; -;
DR EMBL; K03416; AAA50965.1; -;
DR EMBL; K03418; AAA51461.1; -;
DR EMBL; AF109077; AAD26134.1; -;
DR EMBL; AF169078; AAD26132.1; -;

DR EMBL; AF109082; AAD26132.1; JOINED.
DR EMBL; AF109078; AAD26133.1; -;
DR EMBL; AF109084; AAD26133.1; JOINED.
DR EMBL; AF109079; AAD26130.1; -;
DR EMBL; AF109081; AAD26130.1; JOINED.
DR EMBL; AF109079; AAD26131.1; -;
DR EMBL; AF109083; AAD26131.1; JOINED.
DR EMBL; AF109080; AAD26135.1; -;
DR EMBL; AE003454; AAF46732.1; -;
DR EMBL; X02933; CAA26157.1; -;
DR EMBL; X78920; CAA55523.1; -;
DR EMBL; X78918; CAA55521.1; -;
DR EMBL; X78919; CAA55522.1; -;
DR PIR; A00640; GQFFE.
DR HSSP; P11362; 1FGK.
DR FlyBase; FBgn003731; Eogr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR PRINTS; PR001030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW Developmental protein.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 869 889 POTENTIAL.
FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 938 1198 PROTEIN KINASE.
FT NP_BIND 944 952 ATP (BY SIMILARITY).
FT BINDING 971 971 ATP (BY SIMILARITY).
FT ACT_SITE 1063 1063 BY SIMILARITY.
FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

Query Match 28.7%; Score 1964; DB 1; Length 1426;
Best Local Similarity 32.8%; Pred. No. 3e-97;
Matches 470; Conservative 190; Mismatches 422; Indels 352; Gaps 44;

QY 24 QVCTGDMKRLPASPETHLDMLRHLYGQCQVVOGNLELTYP-T-NASLSFLQDIOEVQG 82
DB 100 KICIGTKSRLSPVSKHEHRYNLRDRYNTCTVVDGNLKLTLPLNENLDLSFLDNIREVTG 159
QY 83 YVLIHNVQVQVLPQRLRIVRGTLF-----EDNVALAVLDNGDPLNNTPTVTGASPGGL 137
DB 160 YLILSHVDVKKVFPKQLQIRGTLFSLSVSEKVALFV-----TYSKM 203
QY 138 RELQLRSLEILKGVLIQNRNQLCVQDTILWKDF--NFTVSFW---LRVPKVSASHLEN 193
DB 204 YLEIPDLRLVNGVQGFHNNYLNCHMRTIQSEIVSNGTDAYNVDFTAPE----- 255
QY 194 RSRACHPCSPMKGRCSRGESSEDCQSLSLRTTRVCAGGCA--RCKGPLPTOCCHQCAAGCT 251
DB 256 --RECPKCHESCTHG--CWGEGPKNCQKSKLTCSPQAGGRCYGPKPRECCHLFCAGGCT 312
QY 252 GPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNVYLS 311
DB 313 GPTQKDCIACKNFFDEAVSKCECPMRKYNPTTYVLETNPEGKYAYGATCVKECP--GHLL 371
QY 312 TDVSGSCTLVCPLNQEVTAEDGTQCEKSKCARVCYGLGMEHLREVRVAVTSANTQFEA 371
DB 372 RUNGACVRSRCPQDQKMDKQGE-----CVPNGCPKPTCPGVTVLH-----AGNIDSR 418
QY 372 GCKKIFGSLAFLPESFDG--DPASNTA-----PLQPEQLQVLFETLEITGYLYISAMPD 423


```
Db 118 APNOAHLRLKETEFKVKVLGSGAGFTYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
Qy 772 EILDYAYMAGVSPVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRLGSGDGLLAW 831
Db 178 EILDYAYMAGVSPVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRLGSGDGLLAW 237
Qy 832 CMQIAGMSYLEDLVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGCKVP 891
Db 238 CVQIAGMNYLEERLVRDLAARNVLVKTPOHVKITDFGLAKLGADEKEYHAEGCKVP 297
Qy 892 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQP 357
Qy 952 ICTIDVYIMVKCMWIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYIMVKCMWIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLGPASPLDSTF 417
Qy 1011 YRSLLDEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRRHSSTSRGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEEEAPRSL-----APSEGAGSDVFDGLGMAAGKGLQSLPTHDPSPLOQYSEDP 1125
Db 450 -----SRTPLLSSLSATSNNSATNCID-----RNGQGHVPVREDSFVQYSSDPTGN 495
Qy 1126 LPSET--DGYVAPLTCSPQPEYVQNDVRPQPPSPREGPLPAARPAGATLERAKTSLSPCK 1183
Db 496 FLEESIDDDGL-----PAPEYVQ--LMPKPS-----TAMVQ 526
Qy 1184 NGVVKDVF-----AFGGAVENTPEYLTPOGGAAPQHPHPPAFSPAFDNLXY 1228
Db 527 NOIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPY 578
Qy 1229 WQO-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 579 WIQSGNHQINLONPDYQDQFLPNETKPNGLLKVPAENPEYLRVAAP 625

RESULT 13
ERBB AVIER
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN (2)
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benalissa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
```

```
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02006; AAA42394.1; ALT_INIT.
CC DR EMBL; K01216; AAA42400.1; -.
CC DR PIR; A00644; TVYUH.
CC DR HSP; P11362; LFQK.
CC DR InterPro; IPR000719; Euk_pkinase.
CC DR InterPro; IPR001245; Tyr_pkinase.
CC DR Pfam; PF00069; pkinase; 1.
CC DR ProDom; PD000001; Euk_pkinase; 1.
CC DR SMART; SM00219; TyKc; 1.
CC DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC KW Glycoprotein; Phosphorylation.
CC FT DOMAIN 132 399 PROTEIN KINASE.
CC FT NP_BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC FT ACT_SITE 257 257 BY SIMILARITY.
CC FT CONFLICT 29 29 R -> W (IN REF. 2).
CC FT CONFLICT 140 140 S -> F (IN REF. 2).
CC FT CONFLICT 146 146 I -> V (IN REF. 2).
CC FT SEQUENCE 604 AA; 67633 MW; 76BECDD06745D609 CRC64;
Qy 593 CAHYKDPFCVACPSGVKPDLSYMPKFPDEEGACQPCINCTHSCVDLDDKGPABEQ 652
Db 3 CAHFIDGPHCVKACPAVLGENDTL-VRKYADANAVCQLCHPNCRTGCKGPGLEGCP--- 58
Qy 653 RASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMRLIQETELVEPLTPSGA 711
Db 59 NSGKTPSAAGVVGGLCLVVGGLGLVLR--HIVKRTLRLLQERLEVEPLTPSGE 117
Qy 712 MPNQMRILKETELRKVKVLGSGAGFTYKGLWIPEGEKVKIPVAIKELREATSPKANK 771
Db 118 APNOAHLRLKETEFKVKVLGSGAGFTYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
Qy 772 EILDYAYMAGVSPVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRLGSGDGLLAW 831
Db 178 EILDYAYMAGVSPVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRLGSGDGLLAW 237
Qy 832 CMQIAGMSYLEDLVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGCKVP 891
Db 238 CVQIAGMNYLEERLVRDLAARNVLVKTPOHVKITDFGLAKLGADEKEYHAEGCKVP 297
Qy 892 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQP 357
Qy 952 ICTIDVYIMVKCMWIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYIMVKCMWIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLGPASPLDSTF 417
Qy 1011 YRSLLDEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRRHSSTSRGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEEEAPRSL-----APSEGAGSDVFDGLGMAAGKGLQSLPTHDPSPLOQYSEDP 1125
Db 450 -----SRTPLLSSLSATSNNSATNCID-----RNGQGHVPVREDSFVQYSSDPTGN 495
Qy 1126 LPSET--DGYVAPLTCSPQPEYVQNDVRPQPPSPREGPLPAARPAGATLERAKTSLSPCK 1183
Db 496 FLEESIDDDGL-----PAPEYVQ--LMPKPS-----TAMVQ 526
Qy 1184 NGVVKDVF-----AFGGAVENTPEYLTPOGGAAPQHPHPPAFSPAFDNLXY 1228
Db 527 NOIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPY 578
Qy 1229 WQO-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 579 WIQSGNHQINLONPDYQDQFLPNETKPNGLLKVPAENPEYLRVAAP 625

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 8.7e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-149-163-14
Perfect score: 6848
Sequence: 1 MELAALCRWGLLLALLPPCA.....TFKGTPTAENPEYLGLDVVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 21:*

- 1: sp_archea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 6198 | 90.5 | 1259 | 6 | O18735 |
| 2 | 3166 | 46.2 | 1209 | 11 | Q9QX70 |
| 3 | 3139 | 45.8 | 1210 | 11 | Q9EP98 |
| 4 | 2724 | 39.8 | 1165 | 13 | Q9YH40 |
| 5 | 2711.5 | 39.6 | 1137 | 13 | Q9W6F6 |
| 6 | 2307 | 33.7 | 1328 | 13 | P73754 |
| 7 | 2047.5 | 29.9 | 1433 | 5 | Q9BIH9 |
| 8 | 1777.5 | 26.0 | 419 | 4 | Q9UK79 |
| 9 | 1739 | 25.4 | 367 | 11 | Q8R2X1 |
| 10 | 1720 | 25.1 | 729 | 15 | Q8G712 |
| 11 | 1718 | 25.1 | 567 | 15 | Q8G714 |
| 12 | 1697.5 | 24.8 | 412 | 4 | Q8WYV0 |
| 13 | 1653.5 | 24.1 | 962 | 15 | Q64895 |
| 14 | 1645 | 24.0 | 545 | 15 | Q85468 |
| 15 | 1530.5 | 22.3 | 655 | 11 | Q9WVF5 |
| 16 | 1514.5 | 22.1 | 643 | 11 | Q9ERV6 |

17 1273 18.6 1193 5 Q9YIX8
18 1194.5 17.4 1368 5 Q23821
19 1157 16.9 1717 5 Q26566
20 1142 16.7 527 13 Q90836
21 1025.5 15.0 478 11 Q9SE0
22 952.5 13.9 599 13 Q9PSH2
23 906 13.2 165 4 Q14256
24 887 13.0 176 11 Q923V5
25 806.5 11.8 346 13 P11776
26 778 11.4 435 5 Q8SZW1
27 754.5 11.0 311 13 Q99162
28 749.5 10.9 1362 13 Q9PVZ4
29 725 10.6 1671 5 Q9NJV5
30 722 10.5 149 6 Q9BG66
31 721 10.5 331 4 Q9BUD7
32 702 10.3 1418 13 Q93457
33 699.5 10.2 1368 13 Q8UW85
34 682 10.0 1358 13 Q73798
35 677.5 9.9 1369 13 Q8UW86
36 665.5 9.7 1472 5 Q9U5A8
37 660 9.6 1412 13 Q8UW84
38 647.5 9.5 1418 13 Q8UW83
39 642 9.4 1245 13 Q9YGH8
40 639 9.3 1371 11 Q9QVW4
41 598 8.7 987 11 Q9IYM0
42 597 8.7 935 4 Q96L35
43 595 8.7 987 11 Q99MR2
44 594.5 8.7 2144 5 Q9VD94
45 587.5 8.6 1036 4 Q07912

ALIGNMENTS

RESULT 1

O18735
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB008451; BAA23127.1; -
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TVR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

| Matches | 641; | Conservative | 166; | Mismatches | 353; | Indels | 126; | Gaps | 26; |
|---------|------|---|------|------------|------|--------|------|------|-----|
| Qy | 3 | LAALRCMGLLLALLPPGA--ASTQVCTGTDMLKRLPASPETHLDMRLHLKYQCQVVOGNLE | 61 | | | | | | |
| Db | 15 | LAALCAAG-----GALEEKVKCGTSNRLTQGTGFEDHFLSLQRMFNCEVVLGNLE | 66 | | | | | | |
| Qy | 62 | LTYLPTNASLFLQDIOIEQVOGVYLIAHNQVQFQLRLRIVRGTQLFEDNYALAVLNDGD | 121 | | | | | | |
| Db | 67 | ITYVORNYDLSLFKTIOEVAGVYLIATLNTVERIPLNLQIIRGNALYENTVALAVLSN-- | 124 | | | | | | |
| Qy | 122 | PLNNTTPTVGTASPGGLREQLRLSTEILKGGVLQIRNPOLCYQDTILWKDFNNFTVFWL | 181 | | | | | | |
| Db | 125 | -----YGNKTKGLRELPRNQLQELLIGAVRFSNNPILCNMTIOWRDI----- | 167 | | | | | | |
| Qy | 182 | RVPKVSASHL----ENRSRACHPCSPMCKGSRCSGESDQCSLTRTVCAGGCA-RCKGP | 236 | | | | | | |
| Db | 168 | -VDVFLSNMSMDVQRHLTGCPKCDPSPCNSCWGRGEENCOKLTKIICAQCCSRRCGR | 226 | | | | | | |
| Qy | 237 | LPTDCHEQCAAGCTGPKHSOCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYT | 296 | | | | | | |
| Db | 227 | SPSCCHNQCAAGCTGPRSDCLVCHRPRDEATCKDTPPLMLYNPTTYQMDVNPPEKYS | 286 | | | | | | |
| Qy | 297 | FGASCTVACPYYLSTDVGSCTLVCPHLNQEVTAEDGTQRECKSCSKCARCYCLGMHGL | 356 | | | | | | |
| Db | 287 | FGATCVKCPKNYVYTDHGSVCRAAGPDYIEV-BEDGVSKCKDGGPCRVCNGIGIGE | 345 | | | | | | |
| Qy | 357 | REVRASVTSANTQEFAGCKKIFGSLAFUPESPDGDPASNTAPLOEQLOQVFFELBEITGYL | 416 | | | | | | |
| Db | 346 | KDTLSINATNIKFKYKTAISGDLHLVPAPKGSFTRTPPLDPRELEILTKVKEITGFL | 405 | | | | | | |
| Qy | 417 | YISAWPSLPLDLSYFQNLQVIRGRILUNGAYSLTQGLGISWGLRSLRSLRELGSGLAIHH | 476 | | | | | | |
| Db | 406 | LQAWPENWTDLHAFENLEIIRGTQKHQOPSLAVVGLNLTSLGLRSLKETSDDGVIIISG | 465 | | | | | | |
| Qy | 477 | NTHLCFVHTVPDOLFRNPHQALLHTANRDECECVGSEGLACHOLCARGHGWGPGTPQCVN | 536 | | | | | | |
| Db | 466 | NRNLQYANTINWKKLFGTPNQTKIMNRAEKCKATNHNVCNPLCSSEGGCGPEPTDQVS | 525 | | | | | | |
| Qy | 537 | CSQFLRQGECEECRLVQGLPREYVYNARHCLPHCEQPQONGSVTCFGEPADOCVACAHY | 596 | | | | | | |
| Db | 526 | CONVSRGRECVDKNILEGEPRFVENSECLOCHPECLPTQMTNITCTGRGPDNCKICAHY | 585 | | | | | | |
| Qy | 597 | KDPFCVAPCSPGVKPDLSYMPHWKFPDEGACQPCPNCTHSCVDLDDKGCFAEQRASP | 656 | | | | | | |
| Db | 586 | VDGPHCVKTCPSGIMGENNTL-VNKFADANNVCHLCHANCTYGCAGPLAGC--QQPEGP | 642 | | | | | | |
| Qy | 657 | -LTSIVSAVGLILLAVVLGVVFGI-LIKRQOKIRKYMRLLOETELVPLTPSGAMPN | 714 | | | | | | |
| Db | 643 | KIPSIATGIVGGLLUFIVV-VALGIGLFMRRLQVFKETRLRLQBELVPLTPSGEAPN | 701 | | | | | | |
| Qy | 715 | QAQWRILKETLRKVKVLGSAFTGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIL | 774 | | | | | | |
| Db | 702 | QAHRLILKETFKIKVLGSAFTGVYKGLWIPGEKVKIPVAIKELREATSPKANKEIL | 761 | | | | | | |
| Qy | 775 | DEAYMAGVGSYYVSRLLGLCLTSTVOLVTLQMPYGCLLDHRNRLGSLQDLNWCQ | 834 | | | | | | |
| Db | 762 | DEAYMASVDNPHVCRLLGLCLTSTVOLVTLQMPYGCLLDHYREHKDNIGSQYLLNWCQV | 821 | | | | | | |
| Qy | 835 | IAGMSYLEDLVRLVHRDLAARNVLVSPNHVKITDFGLARLLIDIDEYHADGCKVPIKW | 894 | | | | | | |
| Db | 822 | IAGMNYLLEDRRLVHRDLAARNVLVTPQHVKITDFGLAKLLGAEKEYHAEGGKVPKW | 881 | | | | | | |
| Qy | 895 | MALESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICT | 954 | | | | | | |
| Db | 882 | MALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSTLEKGERLPPPICT | 941 | | | | | | |
| Qy | 955 | IDVYIMVVKCWMIDSECRPRFRELVSFESRMARDPQRVVVIQ-NEDLGPASPLDSTFYRS | 1013 | | | | | | |
| Db | 942 | IDVYIMVVKCWMIDADSRPKFRELILIEFSKARDPQRYLYIQDERMHLPSPTSDSNFYRA | 1001 | | | | | | |
| Qy | 1014 | LLEDDDMGLDVADEEYLVPOOGFFCPDPAFCAGMWHRRHSSTSSGGDLTLGLEPSE | 1073 | | | | | | |
| Db | 1002 | LMEDEEDVDVADEYLVPOOGFF-----NSPST----- | 1030 | | | | | | |

```

RESULT 3
Q9EP98
ID Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Scheel C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAC28045.1; -
DR EMBL; AF275364; AAC28045.1; JOINED.
DR EMBL; AF275365; AAC28045.1; JOINED.
DR EMBL; AF275367; AAC24386.1; -
DR HSP; P11362; IFGK.
DR MGI; MGI:95294; Egr.
DR InterPro; IPR000345; CytC heme bind.
DR InterPro; IPR000494; EGRF_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_chr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.

```

| DR | PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1. |
|-----|--|
| DR | PROSITE; PS00107; PROTEIN KINASE_ATP; 1. |
| DR | PROSITE; PS00011; PROTEIN KINASE_DOM; 1. |
| DR | PROSITE; PS00109; PROTEIN KINASE_TYR; 1. |
| KW | ATP-binding; Receptor; Transferase. |
| SEQ | SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64; |

| Query Match | | 45.8%; Score 3139; DB 11; Length 1210; |
|---|--|--|
| Best Local Similarity | | 49.5%; Pred. No. 5.7e-228; |
| Matches 633; Conservative 170; Mismatches 359; Indels 116; Gaps 233 | | |
| Qy | 11 LLLALLPPGAA--STOVCTGTDMLRLPASPTHLDMLRLHYOGCVQVQGNLELTPLTN 68 | |
| Db | 14 LLTALCAAGGALAEKKVCGQTSNRLTQLGTTFEDHFLSLQRMYNNCCEVGLNLEIITYVQRN 73 | |
| Qy | 69 ASLSFLQDIQEVQGYVLIATHNQVQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNNTTP 128 | |
| Db | 74 YDLSFLTKIQEVAGVYLIALMTVERIPLLENQIIRGNALYENTYALAILSN----- 124 | |
| Qy | 129 VTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDFNFVTSFWLRVPKQVA 188 | |
| Db | 125 -YGTNRITGLRELPMRNLEQELLIGAVRFNSNNPILCNMDTIQRDI-----VQNVMF 173 | |
| Qy | 189 SH-----LENRSRACHPCSPMCKSGRCWGESSEDCSLTRTCVAGGCA-RCKGPLTDCCH 243 | |
| Db | 174 SNMSMDLQSHPSKCPKCDPSPNGSCWGGGENCOKLTKIICAOQCSHRCGRSPSDCCH 233 | |
| Qy | 244 EQCAAGCTGPKHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVT 303 | |
| Db | 234 NOCAAGCTGPRESCLVCKQFDEATCKDTCPPLMLNPTTYQMDVNPBGKYSFGATCVK 293 | |
| Qy | 304 ACPVNYLSTDVGSCTLVCPPLHQEVTAEDGTQRCEKCKPCARVCYGLGMEHLREVRAT 363 | |
| Db | 294 KCPRNVVYVTHGSSVRACGPDYEV-EEGDIRKCKCKDQPCRKVCNGIGIGEFKDKTLSIN 352 | |
| Qy | 364 SANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEITGYLIYSAMPD 423 | |
| Db | 353 ATNFKHKPYCTAISGDHLIPVAFKGSFTPTPLDPRELEILTKVKEITGELLTQAMPD 412 | |
| Qy | 424 SLPDSLVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRELGSGLALIHNNHLCFV 483 | |
| Db | 413 NWTDLHAFENLEIIRGTKHQHQSFLAVVGLNITSLGLRSLKEISDGDVILSGNRLCYA 472 | |
| Qy | 484 HTVPWDQLFRNPHQALLHTANRDEDCVGEGLACHQLCARGHGCGPPTQCVCNCSQFLRG 543 | |
| Db | 473 NTINWKKLFGTPNQKTKIMNNRABKDKAVNHVCNPLCSSEGCWGPEDRCVSCQNVSRG 532 | |
| Qy | 544 QECVEECRVLQGLPREYVYVNRHCLPCHPECPONGSVTCFGEADQCACAHYKDPPEVC 603 | |
| Db | 533 RECVEKNILGEPRFVENSECLOCHPECLPQAMNITCTGRGPONCLOCAHYIDGPHCV 592 | |
| Qy | 604 ARCPSGVVKPDLUSYMPIWKFPPDEEGACQCPINCTHSCVDLDDKCPAEQASPLTSIVSA 663 | |
| Db | 593 KTCFAGIMGENNTL-VMKYADANNVCHLCHANCTYGCAGPGIQCCEVWPSGPKIFSIATG 651 | |
| Qy | 664 VVGLLVVLGVFGEI-LIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAOIRLK 722 | |
| Db | 652 IVGGLLFTIV-VALGIGLPMRRRHIVKRTLRRLQERELVEPLTPSGEADPQAHLRLK 710 | |
| Qy | 723 ETELKVKVILGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 782 | |
| Db | 711 ETEFPKIKVLGSGAFGVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDEAYVMAS 770 | |
| Qy | 783 VGSYPVSRLLGICLTSTVOLVTLQMPYCGLLDHRNENRGRGSDQLLNCWMOIAGKMSYL 842 | |
| Db | 771 VDNPHVCRLLGICLTSTVOLITQLMPYCGLLDYVREHKDNTGSQVLLNCWQVIAKGMNVL 830 | |
| Qy | 843 EDVELVRHDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIKWMALESILR 902 | |
| Db | 831 EDRLEVRHDLAARNVVKTPQHVKITDFGLAKLGAEEKEYHAEGKVPKIKWMALESILH 890 | |
| Qy | 903 RRFTHQSDVASYGVTVHMLTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDYVMIV 962 | |

| | | | | | | | | | |
|----------|--|---|-----------------------------------|------------------------------|----------|--|--|--|--|
| Db | 891 | RIYTHOSDVNSYGVTVWELMTFGSKPYDGI | PASDISISILEKGERLQPPCTIDVYMIW | 950 | | | | | |
| Qy | 963 | KCMIDSECRPREFELVSEFSEMRARDQ | QRFVWIQ-NEDLGPASPLDSTFYRSLLEDDMG | 1021 | | | | | |
| Db | 951 | KCMWIDADSRPKPRELILFEFSKWARD | QRYLVYIQGDERMHLPSPTDSNFYRALMDEDM | 1010 | | | | | |
| Qy | 1022 | DLVDAEYLVPOOGFFCPDPAFCAGAG | VMVHRHRSSTSCGGDLTLGLPSEEEAPRSPL | 1081 | | | | | |
| Db | 1011 | DVYDADEYLTPOQGF | -----NSPST-----SRTPL | 1036 | | | | | |
| Qy | 1082 | APSEGAAGSVDFDGLGMAAKGLQSL | THDPSPLOQRYSEDPTVLPSET--DCGVAPLTC | 1139 | | | | | |
| Db | 1037 | LSLSLATS | -----NSTVACINRNGSCRVKEDAF | LOQRYSSDPTGAVTENDIDDAFL----- | 1087 | | | | |
| Qy | 1140 | SPOPEYVNPQDVRPQPPSPREGPL | PAARPAGATLERAKTLSPOKNGVVKDVAF | GGAVEN | 1199 | | | | |
| Db | 1088 | -PVPEYVNO-SVPKRPAGSVQNPV | YHNQPLHP-----APGRDLHYQN-- | PHSNAVGN | 1135 | | | | |
| Qy | 1200 | PEYL-TPOCGAARQPHPPAFSAF | DNLYYWDQ-----DP----- | PERGAPPST | 1242 | | | | |
| Db | 1136 | PEYLNTAQ | -----PTCLSSGFGNSPALWIKGSH | MSLDNPDYQQDFFPKETKPNGI | 1186 | | | | |
| Qy | 1243 | FKGTPTAENPEYLGLDVP | 1260 | | | | | | |
| Db | 1187 | FKG-PTAENAEYLRVAPP | 1203 | | | | | | |
| RESULT 4 | | | | | | | | | |
| Q9YH40 | | | | | | | | | |
| ID | Q9YH40 | PRELIMINARY; | | PRT; | 1165 AA. | | | | |
| AC | Q9YH40; | | | | | | | | |
| DT | 01-MAY-1999 | (TrEMBLrel. 10, Created) | | | | | | | |
| DT | 01-OCT-2000 | (TrEMBLrel. 15, Last sequence update) | | | | | | | |
| DT | 01-JUN-2002 | (TrEMBLrel. 21, Last annotation update) | | | | | | | |
| DE | Receptor tyrosine kinase proto-oncogene. | | | | | | | | |
| GN | XMRK. | | | | | | | | |
| OS | Xiphophorus xiphidium. | | | | | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | | | | |
| OC | Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; | | | | | | | | |
| OC | Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; | | | | | | | | |
| OC | Cyprinodontiformes; Poeciliidae; Xiphophorus. | | | | | | | | |
| OX | NCBI_TaxID=8086; | | | | | | | | |
| RP | [1] | | | | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | | | | |
| RC | STRAIN=RIO PURIFICATION; | | | | | | | | |
| RC | MEDLINE=98241172; PubMed=9582016; | | | | | | | | |
| RX | Dimirijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J., | | | | | | | | |
| RA | Altschmid J., Scharltl M.; | | | | | | | | |
| RT | "Activation of the Xmrk proto-oncogene of Xiphophorus by | | | | | | | | |
| RL | overexpression and mutational alterations."; | | | | | | | | |
| RL | Oncogene 16:1681-1690(1998). | | | | | | | | |
| RN | [2] | | | | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | | | | |
| RC | STRAIN=RIO PURIFICATION; | | | | | | | | |
| RA | Scharltl M.; | | | | | | | | |
| RA | Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases. | | | | | | | | |
| DR | EMBL; U53471; AAD10500.2; -- | | | | | | | | |
| DR | HSSP; P11362; 1FGK | | | | | | | | |
| DR | InterPro; IPR000345; CytC_heme_bind. | | | | | | | | |
| DR | InterPro; IPR000494; EGFR_L_domain. | | | | | | | | |
| DR | InterPro; IPR000719; Euk_pkinase. | | | | | | | | |
| DR | InterPro; IPR002174; Furin-like. | | | | | | | | |
| DR | InterPro; IPR001899; Gram_pos_anchor. | | | | | | | | |
| DR | InterPro; IPR001245; Tyr_pkinase. | | | | | | | | |
| DR | Pfam; PF00757; Furin-like; 1. | | | | | | | | |
| DR | Pfam; PF00069; pkinase; 1. | | | | | | | | |
| DR | Pfam; PF01030; Recep_L_domain; 2. | | | | | | | | |
| DR | PRINTS; PR00109; TYRKINASE. | | | | | | | | |
| DR | ProDom; PD000001; Euk_pkinase; 1. | | | | | | | | |
| DR | SMART; SM00261; Fur; 3. | | | | | | | | |
| DR | SMART; SM00219; TyrKc; 1. | | | | | | | | |
| DR | PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1. | | | | | | | | |

| | | | |
|---|---|--|---------------|
| Best Local Similarity 46.9%; Pred. No. 1e-195; | | Matches 540; Conservative 170; Mismatches 346; Indels 95; Gaps 27; | |
| Qy | 161 | LCVQDTLWKDF-----NNTVSWLRVVKVSASHLENRSRACHPCSPCKGSRGWGES | 214 |
| Db | 3 | LCFADTTHWQDIVRNWASNFTL-----VP-----TNGSSGCGCRKSKCTG-RCWGPT | 49 |
| Qy | 215 | SEDCOSLTRVCAGGC-ARCKGPLPTCCHEQCAAGCTGPKHSDCLACLPHNHSIGICELH | 273 |
| Db | 50 | ENHCOTLTKVCAEQDGRGVYVSDCHREAGGSGPKDTCFACMFNDSGACVTO | 109 |
| Qy | 274 | CPALVYNTDTFESMPNPEGRTFGASCVTACPNYLSTDVGSCTLCPLHNQEVTAEDG | 333 |
| Db | 110 | CPQTFVYNTPTFQLEHNHNAKYTYGAFCKVKCPNFV-VDSSSCVRACPPSSKMEV-EE | 167 |
| Qy | 334 | TORCEKSKCARVCYGLGMEHLREVRVAVTSANIEFAGCKKIEGSLAFLPESFDGPAS | 393 |
| Db | 168 | IKMKCPCTDIPKACDGGTGSLSVAQTVDSNIDKFINCTKINGNLIFLVTGTHGDPYH | 227 |
| Qy | 394 | NTAPLOEQVFTLEITGVLISAWPSLPLSVFONLQVIRGRILHNGAYSLSLQOG | 453 |
| Db | 228 | TIAAINPEKLNIFQTVREITGVLNQSWPENWTDPRVFSNLVTIGGRALVSGLSLLIKQ | 287 |
| Qy | 454 | LGISWLGSLRELGSLALIHNNHLCFVHTVPMQDLFRNPQOALLHTANRPEDECVGE | 513 |
| Db | 288 | QGITSLOFQSLKQISAGNIVITDNLNLCYVHTVNTSLFSTPSQKTVIHRNKAENCTAD | 347 |
| Qy | 514 | GLACHQICARGHCWGPPTQCVNCSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPEC | 573 |
| Db | 348 | GMVNELCSSDGCWGPDOCLSCRFIRGTCIESCNLYDGEFREFANGSVCMEDCPQC | 407 |
| Qy | 574 | QP-ONGSVTCFGEADOCVACAHYKDPFPCVAPCGVGPVDPKLSYMPILWKPDERGACQC | 632 |
| Db | 408 | EKVEDNMITYCGPDPDCTKCFHFKDPGNCVEKCPDLOGANSF--IFKVADEDRCHPC | 465 |
| Qy | 633 | PINCHSCVLDLDDKGC-----PAEQRASPLTSIVSAVY-GILLVVLGVVFGIL | 680 |
| Db | 466 | HPNCTQCGRPASHDCIYVWPTRQSTLPQHAR-TPL--IAAGVIGGLFIIVIMGLTFAVY | 522 |
| Qy | 681 | IKRQOKIRKYWRRLQETELVEPLTPSGMNPQAOQWRLKETELRKVKVLSGAGTV | 740 |
| Db | 523 | VRKRSIK-KKRALRRFL-ETELVEPLTPSGTAPNQAOLRIKETELRKVKVLSGAGTV | 580 |
| Qy | 741 | YKGIWIPDGENVKIPVAIKVIRENTSPKANKEIIDEAYVWAGVSPVYVSRLLGLCLSTV | 800 |
| Db | 581 | YKGIWVPEGETVKIPVAIKILNETTGPKANVEFNDEALIMASMDPHVRLVGLCLSPFI | 640 |
| Qy | 801 | QLVTQLMPYGLLDHVRNRRGLSGQDLLNMCQIAKMSYLEDLVRLVHRDLAARNVLVK | 860 |
| Db | 641 | QLVTQLMPHGLLDYVHEHKDNIQSOLLNMCVQIAKGMYLEERLVRHDLAARNVLVK | 700 |
| Qy | 861 | SPNHVKITDFGLARLLDIDETEXHADGKVPKIMWALESILRRBFTHQSDVMSYGVTVWE | 920 |
| Db | 701 | SPNHVKITDFGLARLLSGDEKEYNADGGKMPKIMWALECIHYRKFTHQSDVMSYGVTVWE | 760 |
| Qy | 921 | LMTFGAPYDGPAREIPDLLEKGERLPQPPICTIDVYMWKCMWTDSECRPRFELVS | 980 |
| Db | 761 | LMTFGGPDYDGPITREIPDLLEKGERLPQPPICTIDVYMWKCMWTDASRPKELAA | 820 |
| Qy | 981 | EFMRMARDPQRFVVIQNEED-LGPASPLDSTFYRSLLEDDDMGDLVDAAEYLVPOQGFPC | 1039 |
| Db | 821 | EFMRMARDPQRYLVIOGDDRMKPLSPNDSKFFQNLDEEDLMDMAEYLVLP-QAFNIP | 879 |
| Qy | 1040 | DPAPGAGMWHHRHSSTSGGDLTLGLEPSEEAPRS--PLAP-SEGAGSDVFDGDL | 1096 |
| Db | 880 | PIYTSRTRIDSNRNQFYRDGGVAABGVV-PMPYRAPGCIIPAPVAQGATAIFEDTC | 938 |
| Qy | 1097 | GMGAAGLQSLTHDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQEVYNQ | 1149 |
| Db | 939 | CNGTLRKQVATLAKEDSSQRYSDPTVFIPERVIRGELEDGMYTMRDKPKTDYLNVP | 998 |
| Qy | 1150 | DVTRPOPPSPREGPLPAA-RPAGATLERAKTSLSPKNGVVKDVF-----AFGGAENP | 1200 |
| | | | |
| Db | 999 | EENPFVSRKXNGDLQAVDNPEYHN-----APNGQPKAEDYVNEPLYLNTFANTLENA | 1051 |
| Qy | 1201 | EYLTPQCGAARPOHPHPAPAFDNLVYWDODPPERGA--PPSTFKGTPT----- | 1248 |
| Db | 1052 | EYL-----KNNLPEKAKAFNDPDYMNHSLPPRSTLQHPDYLQBYSTKYFYKONGRI | 1103 |
| Qy | 1249 | -----AENPEYL 1255 | |
| Db | 1104 | RPIVAENPEYL 1114 | |
| RESULT 6 | | | |
| ID | P79754 | PRELIMINARY; | PRT; 1328 AA. |
| AC | P79754; | | |
| DT | 01-MAY-1997 | (TrEMBLrel. 03, Created) | |
| ..T | 01-MAY-1997 | (TrEMBLrel. 03, Last sequence update) | |
| DT | 01-JUN-2002 | (TrEMBLrel. 21, Last annotation update) | |
| DE | Erbb3. | | |
| OS | Fugu rubripes (Japanese pufferfish) (Takifugu rubripes). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; | | |
| OC | Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; | | |
| OC | Tetraodontidae; Takifugu. | | |
| OX | NCBI_TaxID=31033; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=99177347; PubMed=10077531; | | |
| RA | Gelliner K., Brenner S.; | | |
| RT | "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu | | |
| RT | rubripes"; | | |
| RL | Genome Res. 9:251-258(1999). | | |
| DR | EMBL; AF056116; AAC34391.1; .. | | |
| DR | HSSP; P11362; IFGK. | | |
| DR | InterPro; IPR000494; EGFR_L_domain. | | |
| DR | InterPro; IPR000719; Euk_pkinase. | | |
| DR | InterPro; IPR002174; Furin-like. | | |
| DR | InterPro; IPR001245; Tyr_pkinase. | | |
| DR | Pfam; PF00757; Furin-like; 1. | | |
| DR | Pfam; PF00069; pkinase; 1. | | |
| DR | Pfam; PF01030; Recep_L_domain; 2. | | |
| DR | ProDom; PD000001; Euk_pkinase; 1. | | |
| DR | SMART; SM00261; FU; 3. | | |
| DR | SMART; SM00219; TyKc; 1. | | |
| DR | PROSITE; PS00107; PROTEIN_KINASE_ATP; 1. | | |
| DR | PROSITE; PS00101; PROTEIN_KINASE_DOM; 1. | | |
| DR | ATP-binding; Transferase. | | |
| KW | ATP-binding; Transferase. | | |
| SQ | SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64; | | |
| Query Match 33.7%; Score 2307; DB 13; Length 1328; | | | |
| Best Local Similarity 40.5%; Pred. No. 4.6e-165; | | | |
| Matches 525; Conservative 152; Mismatches 414; Indels 206; Gaps 33; | | | |
| Qy | 9 | WGULLALLPP--GAASTQ-----VCTGTDMLRLPASPETHLDMRLHLYQGCVVQGNLEL | 62 |
| Db | 4 | WRLILMCVSRRLRAASSQTQEAQVCPGTQNGSLSTGSEQNYLNKDRYKGEIINGNLEI | 63 |
| Qy | 63 | TYLPTNASLSFLQDIQEVQVYLIAHQVQVLPQRLIRVGTOLFDENYALAVLDNGDP | 122 |
| Db | 64 | QTIESNWDFFSLKTIREVTGVYLIAMNHQEIPLGQLQIRVIRGNSLYERRFALSVELN--- | 120 |
| Qy | 123 | LNNTTPTVGTASPGRLRELQRLSLTEILKGGVLQIRNPOLQCYQDTILWKDF--NN---FTV | 177 |
| Db | 121 | -----YPKDG--PSGLNQLGLMNLTEILDGGVQVINNKYLRYGPPVYWRDIIRNDAPTEI | 174 |
| Qy | 178 | SFMLRVPKVSASHLENRSRACHPCSPMKGSRGSESSDCQSLTRTVCAAGC--ARCKGP | 236 |
| Db | 175 | QF-----NGERG--CHKSC-GNYCWGPGKQCCQLTKTVCAPOCNDRCFGT | 218 |
| Qy | 237 | LPTDCCHQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRT | 296 |
| Db | 219 | SPRDCCHIECAAGCKGPLDTCFACRLFNDSGACVPOCPQTLIYNKQTFQMETPNNAKYQ | 278 |

```
QY 297 FGASCVTACPNYVLSTDVSGCTLVCPHNOEVTAEQGTOR-CEKSKPKCARVCYGLGMEH 355
Db 279 YGSICVSOQCTHFV-VDGSVCVCPDKNHEV--ERGSQRCCLSGCLCPKVCCEGTGAE- 334
QY 356 LREVRVAVTSANTQEAAGCKKIIFGSLAFLPESFDGDPASNTAPLOPEOLOVFEETLEBITGY 415
Db 335 --QRQTVDSNIDSFNCTKIQSLHFLVTGILGDDFKVPPDADKALEVFRVREITDI 392
QY 416 LVISAWPDLSPDLVFQNLQVIRGRILHNGAVSLTLQGLGISWLGSLRSLRELGSLGALIH 475
Db 393 LNIQSPKELNDLSVFSLTTIQGRSLFKRSLFSLVMVRIPTLTLGLSLRSLREISDGSVYIS 452
QY 476 HNTILCFVHTVNDQLFNPH-QALLHTANRDECEVGEGLACHQIACRGHCWGPPQTC 534
Db 453 QNAHLCHYHTVNTQTLFRGSRVANSLSNRMAECVADGRVCDPLCSDSGCWGPPDQC 512
QY 535 VNCSPFLRGOECVEECRLQGLPREYVNAH-CLPCHPECOPONGSVTCFGEADOCVAC 593
Db 513 LSCRYSRHGTCVAGCHFNSGIPREPAGLNGVCVACHPECKPQTGKASTGPGADCEMAC 572
QY 594 AHYKDPFFCVARCPGKVPDLSPYIWKFPDDEGACQPCPINCTHSCVDLDDKGCPAEQR 653
Db 573 TKFRDGPYCMSSCPAGVN-DGEKGLIFKFPNREGHCPCQNCQTCGSGPGLNDC--LE 628
QY 654 ASPLTSIVSAVVGILLVVLGVVF-----GLIKRROOKIRKYTMRLLOQETELVEPL 706
Db 629 AARLTSSQITGIALGVPAIGLFLVFLGLMGLYHRGLAIRKRAMRRYLSGESFEPL 688
QY 707 TPGAMPNQAOMRILKETELRKYVKGAFGVYKGIWIPDGENVKIPVAIKVLRENTS 766
Db 689 GP-GEKGTVKHARILKPSDLRKIKPLGSGVGTGSKGFWIPGEETVKIPVAIKTQDSSG 747
QY 767 PKANKEILDEAYMAGVGSYVSRLLGICLTSTVQLVTOLMPYGCILLDHVRENRLGSG 826
Db 748 RQTFTETIDHLLSMGSLDHPYVRLIGICPGCLQLVTQLSSGSLLEHIRQHKTSLDQ 807
QY 827 DLLNMCQIAKMSVLEDLVLRDLAARNVLKSNHVKIYDFGLARLLDIDETEHAD 886
Db 808 RLNNVCQIAKMYLEEHVRVHKNLAARNILLKNDYQVOISDYGVADLLYPDDKKYVYS 867
QY 887 CGKVPKIMWALESIILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGR 946
Db 868 ETKTKPIKMALESILFRYTHQSDVMSYGVTVWEMMSFGAEPIASVQPEVSVLEKGR 927
QY 947 LPQPFITIDVYMWKCMIDSECRPRELVSEFSRMDPQRPFVVIQNEDLGPASPL 1006
Db 928 LSQPAICTIDVYMWKCMIDENIRPTFKELASDFTRMARDPPRYLVIRMEG----- 980
QY 1007 DSTFYRSLLEDDMDGLVDAAEYLVPOQGFPCDPAPGAGMVHHRSSSTRSGGDLT 1066
Db 981 -----EDSGMGEFL-----RAGSER--GLLE 999
QY 1067 LCLEPSEBEAPRSLAPSEGAGSDVFDGLGMG---AAKGLQSLPHDPSPLQ----- 1116
Db 1000 ADLEDEEB-----GLGDRFATPSLQPSFSWSTSPSQINSYMW 1038
QY 1117 ---RVSEDPVLPSETGYVAPLTCSPQ- EYVQ-----PDVRPQPPSP 1158
Db 1039 TQLRYD-----FAVSQGHIGYLPWSPSPVDITRQLWYQSRSLSSVRLTPDRSAFRSS 1092
QY 1159 REGPL- PAARPAGATLERAKTLSPGKGVKDVFAFGGAVENPEYLPQGGAAAPQHPH 1216
Db 1093 REAECEGAQAGIFRV-----FGSERGN-----PQGG----- 1122
QY 1217 PAFSPAFLNYWDQDPPPERGAPPSTFKGTPTAENPE 1253
Db 1123 -----QQRKLSTASSPSSFKTWADEEDE 1146

RESULT 7
Q9BIH9 PRELIMINARY: PRT: 1433 AA.
AC Q9BIH9;
```

```
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]_TaxID=7165;
RP SEQUENCE FROM N.A.
RC STRAIN=SUA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC heme bind.
DR InterPro; IPR000494; EGFR L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER.
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.9%; Score 2047.5; DB 5; Length 1433;
Best Local Similarity 32.7%; Pred. No. 2e-145;
Matches 476; Conservative 195; Mismatches 389; Indels 395; Gaps 37;

QY 26 CTGTDMLKRLRASPETHLDMLRHLVGGCQVVOGNLELTLYPTVASLSFLQDIOEVQGYVL 85
Db 1 CIGTNGRMSVPANREYHYKNLRDRYTNCTYVDGNLEITWQNIITDLNFIQHIREVYGYVL 60
QY 86 TAAHNOVRQVPLQRLIRVGTQLF-----EDNYALAVLDNGDPLNNTPTVTGASPGGLREL 140
Db 61 ISLYLDLPQVILPRLQIIRGRTTFKLKWEAYGLFV-----SFSHMNTL 104
QY 141 QLRSLTEILKGGVLIQRIQNPOLCYQDTILWKQDNFTVSVFWRVPKVSASHLENRS---RA 197
Db 105 ELPALRDILGSGVGFNNYLNLCMKSIWEEI-----LLAPQTSMQYTFNFSSPERV 156
QY 198 CHPCSPMKGSRCEWGESSEDCQSLRTRTVACGCA--CKGPLPTDCCCHQCAAGCTGPKH 255
Db 157 CPPCHPSCEVG--CWGEGAHNCQRFKJNCSPQSCQRCFCGPKRBCCHFLCAGGCTGPTQ 215
QY 256 SDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYLSLTDVG 315
Db 216 SDCLACKNFDYDGVCKQCEPPMQIYNPTNYFWRPNPDGKYAGATCVKRCP-EHLLKONG 274
QY 316 SCTLVCPHNOEVTAEQGTORCEKSKPKCARVCYGLGMEHLREVRVAVTSANTQEAAGCKK 375
Db 275 ACVRKCPKGMKPNSE-----CVPCKGVCPKTCPEGIVH-----SDNIGNYKDCIT 321
QY 376 IFGSLAFLPESFDGDPASNT-----APLOPEOLOVFEETLEBITGYIYISAWPDLSPD 427
Db 322 IEGSLEILDQSGDFGQQVYTNFSGPRYIKIDPDRLEVPSTVKETITGTFINIQAHHPNFTT 381
```

```
Qy 428 LSVFQNLQVIRGRIHNGAY-SLTLOGLGISWLGRLSRLGSLGLALIHNTHLCFVHTV 486
Db 382 LNTFRNLEVGGRLKENLFASVTIVKTSLSLEKSLKRNSGSIYLENSDLCFVEDI 441
Qy 487 PWOLFNPQHALLHTANRPEDECVBEGGLACHQLCARGHCWGPGPTQCVNCSQFLQOEC 546
Db 442 DWSEIKKSSDHEVMVQKNRNATECHEGEMECSEQCSKAGCWGKGPEQCLECKNVKYGKC 501
Qy 547 VEECRVLQGLPREY-VNARHCLPCHPECOPONGSVTCFGEADOCVACAHYKDPFPCVAR 605
Db 502 LQSCK-SUPRLYSVDSKTCGDCHQCKD-----FCYGNEDNCGSMNVKQGRFCVAE 553
Qy 606 CP-----SGVRPDLSSYPMIFKPPD 624
Db 554 CPTTKHAWGTCINCHKTCVCGCRPRDTIAPDGCISCDAKILGSDAKIERCLMKDSCPD 613
Qy 625 -----BEG----- 627
Db 614 GYSDYVLQEBGLQLSGKAVCRKCHPRCKKCTGYGFHEQFCQECTGYKKGECQECBCP 673
Qy 628 -----ACQCPINCT-----HSCVDL-----DD-----KCCPAEQ- 652
Db 674 QDFYANETRCLPCHOEGRCHGCHGLGDDHCECRNLKLFEGDPYDNATFTTCVSNCPASHP 733
Qy 653 -----RASPLTSIVSAVVGILLVVVLGVVFGI---LIXRQ 685
Db 734 YKRFPPQAGKIGPYCSADSMQSGRIEIPQTVKIVMSGVMALILLCVVGIAFLVFRHK 793
Qy 686 QKIRKYTMRLLQETELVEPLTPSGAMPNQMRILKETELRKVKVLGSGAGFTVYKGIW 745
Db 794 NKDAVKMTALACGEDSEPLRPSNVGNPLTKRIKEAEIRRGVGLGMGAFGRVPGVM 853
Qy 746 IPOGENVKIPVAIKVLRENTSPKANKELLDEAYVMAGVSPYVSRLLGICLTSVQLVTQ 805
Db 854 MPEGESVKIPVAIKVLMEMSGSESKFLEAYIMASVEHFNLLKLAVCMTSQMMLITQ 913
Qy 806 LMPYGCLLDVRNENRGLSDLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHV 865
Db 914 LMPGLCLLDVYRNKKIGSKALLNWSQIARGNAYLEERLVRDLAARNVLVOTPSCV 973
Qy 866 KITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRPTHQSDVMSYGVYVWELMTFG 925
Db 974 KITVFLGAKLLDFDSBYRAAGGKMPKWLALCEIRHRVFTSKSDVMVAFGITIWEELTYG 1033
Qy 926 AKPDGIPAREIPDLLEGERLPPOPCTIDVYIMVVKWIMIDSECRPRELSEFSRM 985
Db 1034 ARPENYPAKDVPPELIEIGHKLQPDICSLDVCILSCWVLDADARPTFKQLAETFAEK 1093
Qy 986 ARDPQRFVIONEDLGPASPLDSTFYRSLLEDDDDMGDLV----- 1024
Db 1094 ARDPGRYLM-----PGDKFMRUPSYTNQDEKDLIRLAPVMAAAAAAAGASN 1144
Qy 1025 -----DAEYLVPOQGFPCPDAPGAGGMVHRHRSSTRSGGDLTLGLLEPSEEA 1076
Db 1145 VDVPSITAEDEVLPQKTRPSIMLPGPSA-----VEPS-DEM 1180
Qy 1077 PRS-----PLAP-----SEGAGSDVPDGLNGAAKGLQSLPTHDPSPLORYSEDPVLP 1127
Db 1181 PKSLRYCKDLPKPDDETDHGKEV-----GVGGIR-----LNLP 1214
Qy 1128 SETDGYVAPLTCSPQPEYVNPQVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGW 1187
Db 1215 LDEDDYLP-TCQSQ-----NQS-----TFG-----Y 1235
Qy 1188 KDVFAFGAVENPYL-----TFQGGAAPOHPHPPPAFSPAFDNLVYWDQPPPERGA 1238
Db 1236 MDLIGVPASVDNPEYLMGSTQAIAGLAQSGM--PHTPP----- 1272
Qy 1239 PPSTFKGTPTAENPE 1253
Db 1273 PPNTPNGMPTHQHSQ 1287
```

```
RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF177761; AAD56009.2; -.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recept_L domain; 1.
DR SMART: SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match .26.0%; Score 1777.5; DB 4; Length 419;
Best Local Similarity 84.2%; Pred. No. 8.1e-126;
Matches 340; Conservative 11; Mismatches 32; Indels 21; Gaps 5;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLYOGCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLYOGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEQVGYVLIAHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEQVGYVLIAHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTCASPGGLRELQLSLTEILKGGVLIQRNPOLCYQDTILWKDF---NNPTV 177
Db 121 DPLNNTTPVTCASPGGLRELQLSLTEILKGGVLIQRNPOLCYQDTILWKDF--- 177
Qy 178 SFWLRVPKVSASHLE-NRSRACHPCSPMKGSCWGESSEDCQSLTRTVCAAGCARCKGP 236
Db 178 -----QLALTLLDTRSRACHPCSPMKGSCWGESSEDCQSLTRTVCAAGCARCKGP 230
Qy 237 LPTDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
Qy 297 FGASCVTACPNYILSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMBHL 356
Db 291 FGASCVTACPNYILSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCAR-----GTHSL 345
Qy 357 REVRAVTSANTIOERFAGCKKIFGSLAFLPESFDGDPASNTAPLQ 400
Db 346 PPRPAVVPVPIRMQPG--PAHPVLFLRPSWDLVSIFYSLPLAP 387

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
```

```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.; Jansson L.; Doederlein G., Lhotak V.,
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027080; AAH27080.1; -.
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match      25.4%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 5.4e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLEKGERLPQPPICT 954
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLEKGERLPQPPICT 60

QY 955 IDVYMTMVKWMIDSECRPRFRELVSFESRMARDPQRFVVIQNEIDLGASPLDSTFYRSL 1014
DB 61 IDVYMTMVKWMIDSECRPRFRELVSFESRMARDPQRFVVIQNEIDLGASPLDSTFYRSL 120

QY 1015 LEDDDMGDLVDAEYLVPOGGFPCDPAPGACGGMVHRSSSTRSGGGDLTLGLEPSEE 1074
DB 121 LEDDDMGDLVDAEYLVPOGGFPCDPAPGACGGMVHRSSSTRSGGGDLTLGLEPSEE 180

QY 1075 EAPRSLAPSEAGSDVDFDGLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGYV 1134
DB 181 EAPRSLAPSEAGSDVDFDGLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGYV 240

QY 1135 APLTCSPOEYVNOQVDRPQPPREGPLPAAPAGATLERAKTLPFGKNGVVKVFAFG 1194
DB 241 APLTCSPOEYVNOQVDRPQPPREGPLPAAPAGATLERAKTLPFGKNGVVKVFAFG 300

QY 1195 GAVENPEYLTQGGAPQHPHAPFAFDNLYWDDPDERCAPSTFKGPTAENPEY 1254
DB 301 GAVENPEYLTQGGAPQHPHAPFAFDNLYWDDPDERCAPSTFKGPTAENPEY 360

QY 1255 LGLDVPV 1261
DB 361 LGLDVPV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities."
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSP; P03322; IAGS.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.

DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;

Query Match      25.1%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 4.1e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 575 PONGSVTCFGEADQCACAHYKDPFCVARCSPGKVPDLSYMPIWKFPEEGACQPCPI 634
DB 141 PEETATPKTGP--DHCWKCAHFIDGPHCVKACPAVGLGENDTL-VMKYADANAVCOLCHP 197

QY 635 NCTHSCVLDLDDKGCAPAEQASPLTSIVSAV--GILLVVVLGVVFGILIKERQOKIRKVTM 693
DB 198 NCTRCCKGCGLEGP---NGSKTPSTAAGVVGGLLCLLVVGLGIGLYLRRR-HIVRKRTL 253

QY 694 RRLQETELVEPLTPSGAMPNQAMRIKTELKRVKVLGSGAGFTVYKGIWIPDGENVK 753
DB 254 RRLQERELVEPLTPSGEAPNQAHILAKETEFKVKVLGSGAGFTVYKGLWIPGSKVK 313

QY 754 IPVAIKVLENTSPKANKELIDWAYVMAGVSGSPYVSRLLGICLTSTVQLVTQMPYGLL 813
DB 314 IPVAIKELREATSPKANKELIDWAYVMASVDNPRVCRLLGICLTSTVQLITQMPYGLL 373

QY 814 DHVRENRLGSGDILLNWCQAKNSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 873
DB 374 DYIREHKDNIGSQYLLNWCQAKNSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 433

QY 874 RLIDDETHYHADGGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGP 933
DB 434 KLLGADEKHYHAGGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGP 493

QY 934 AREIPDLLKGERLPQPICTIDVYIMVYKWMIDSECRPRFRELVSFESRMARDPQRFV 993
DB 494 ASEISVLEKGERLPQPICTIDVYIMVYKWMIDSECRPRFRELVSFESRMARDPQRFV 553

QY 994 VIQ-NEDLGASPLDSTFYRSLLEDMDGLVDAEYLVPOGGFPCDPAPGAGGMVHR 1052
DB 554 VIQGERMHLPSPTDSKFTYRLMEEDMEDIVDAEYLVPHOGFF----- 598

QY 1053 HRSSSTRSGGDLTLGLEPSEEAAPRSL-----APSEAGSDVDFDGLGMAAGLQSL 1107
DB 599 -NSPST-----SPTLLSSLSATSNNSATNCID-----RNCQGH 631

QY 1108 PTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQYVNOQVDRPQPPSPREGPLPA 1165
DB 632 PVREDSFVQRYSSDPTGNFLESIDDGFL-----PAPEYVNO--LMPKKPS----- 675

QY 1166 ARPAGATLERAKTLPFGKNGVVKVDFP-----AFGGAVENPEYL 1203
DB 676 -----TAMVQNOIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL 715

RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
```

RA Johnsson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725B1 CRC64;

Query Match 25.1%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 4e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 584 GPEADQCACAHYKDPFCVARGSPGVKPDLSYMPIWKFPDEEGACQPCPINTHSCVDL 643
DB 1 GP--DHCMKCAHFTDGHPCVKACPAVLGENDTL-VMKYADANAVCQLCHPNCCTRGCKGP 57

QY 644 DDKGCPAEQASPLTSIVSAV-GILLVVLGVVFGILLIKRQOKIRKYTRRLLOTEL 702
DB 58 GLEGG---NGSKTPSAAAGVWGLCLVGVGLIGLYLRR-HIVKRTLRLLQREL 113

QY 703 VEPLTPSGAMPNQAQMIKTELKRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVL 762
DB 114 VEPLTPSGEAPNQAHLRLAETEFKVKVLGSGAFGVYKGLWIPGEKVKIPVAIKEL 173

QY 763 ENTSPKANKIILDEAYVMAGVGSYVSRLLGICLTSTVQLVTLQMPYGCILLDHVRENR 822
DB 174 EATSPKANKIILDEAYVMASVDNPRVCRLLGICLTSTVQLITQLMPYGCILLDVI 233

QY 823 LGSODLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
DB 234 IGSQYLLNMCVQIAKNVYLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGADKE 293

QY 883 YHADGGKVPKIKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDL 942
DB 294 YHAEGGKVPKIKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353

QY 943 KGERLPQPICTIDVYIMVKCWMIDSECRPRFRELVSFMRARDPQRFVVIQ-NEDLG 1001
DB 354 KGERLPQPICTIDVYIMVKCWMIDADSRPKRELIAEFSKWARDPPRYLVIQGDERMH 413

QY 1002 PASPLDSTFVRSLEDDMDGLVDABEYLVPOQGFCCPDAPAGAGGVHRRSSSTRSG 1061
DB 414 LPSPDTSKFRVTLMEEDMDIVDABYLVPHQGF-----NSPST--- 454

QY 1062 GGOLTLGLEPSEEAAPSPL-----APSEAGSGSVDFDGLCMGAAGLQSLPHDPSPLQ 1116
DB 455 -----SRTPLLSLSLATSNNATNCID-----RNGQGHVPVREDSFVQ 491

QY 1117 RYSEDPVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARAGATLE 1174
DB 492 RYSDPTGNFLEESIDGFL-----PAPEYVNO--LMPKKPS----- 526

QY 1175 RAKTLSPGKGVKDVFE-----AFGGAVENPEYL 1203
DB 527 -----TAMVQNOIYNNISLTAISKLPMSRYSQNSHSTAVDNPYL 566

RESULT 12
Q8WYV0
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN Hypothetical 44.7 kDa protein.
GN PF3659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; EF HAND; UNKNOWN 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.8%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 8.8e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 895 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 954
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 60

QY 955 IDVYIMVKCWMIDSECRPRFRELVSFMRARDPQRFVVIQEDLGPASPLDSTFYRSL 1014
DB 61 IDVYIMVKCWMIDSECRPRFRELVSFMRARDPQRFVVIQEDLGPASPLDSTFYRSL 120

QY 1015 LEDDDMGDLVDABEYLVPOQGFCCPDAPAGAGGVHRRSSSTRSGGDLTLGLEPSEE 1074
DB 121 LEDDDMGDLVDABEYLVPOQGFCCPDAPAGAGGVHRRSSSTRSGGDLTLGLEPSEE 180

QY 1075 EAPRSLAPSEAGSGSVDFDGLCMGAAGLQSLPHDPSPLQRYSEDPVPLPSETDGYV 1134
DB 181 EAPRSLAPSEAGSGSVDFDGLCMGAAGLQSLPHDPSPLQRYSEDPVPLPSETDGYV 240

QY 1135 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTLSPGKGVKDVFAFG 1194
DB 241 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTLSPGKGVKDVFAFG 300

QY 1195 GAVENPEYLTPOGGAAPQ-----HPPPA---FSPAFDNL 1226
DB 301 GAVENPEYLTPOGGAALSPDLLLSAQPSSTTIGTRTHQSGGLHPAPSKGHLRQRTQST 360

QY 1227 YYWD-QDPPR-----GAPPSFGKTPTAEN 1251
DB 361 WYWTQCEPEQGVRRSPDVSSSGSREGLTSAGIKRWEGPPTTSRGTCCHARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN GAG, v-ERB-A, v-ERB-B.
OS Avian erythroblastosis virus.

OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruckin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B";
RL Oncogene 5:15-24(1990).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CA936459.1; -
DR EMBL; X52211; CA936459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Steroid_receptor.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; Tyrc; 1.
DR SMART; SM00199; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 24.1%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 6.5e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 547 VEECRVLQGLPRE-VYNAR-HCLP-----CHPEQ 574
DB 354 IEKQESYLAFAYEINVRKHNIPHFWSKLLMKVADLRMIGAYHASFLLHMKVECPTELS 413
QY 575 PQNGSVTCFGEADOCVACAHYKDPFCVAPCPGSKVPLSYMPIWKFPDESGACOPCP 634
DB 414 PQE-----VGP--DHCKMCAHFIDGPHCVKACPAVLGENDTL-VKDYADANAVCOLCHP 465
QY 635 NCTHSCVDLDDKGCAPAEORASPLTSIVSAVV-GILLVVLGVVFGILIKRROOKIRKVTM 693
DB 466 NCTRCCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVGIGLYLRRR-HIVRKRTL 521
QY 694 RLLQETELVEPLTPSGAMPNQAQRILKETELRKVKVLGSGAFGTGVYKGIWIPGENVK 753
DB 522 RLLQERELVEPLTPSGEAPNQAHLRIKETEFKKVKVLGFGAFGTGVYKGLWIPEGEKVT 581
QY 754 IPVALIKVLENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCLL 813
DB 582 IPVALIKVLENTSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLVTOLMPYGCLL 641
QY 814 DHVRENRRGLSQDILLNMCQIAGKMSYLEDVRLVHRDLAARNVLVKSPPNHVKITDFGLA 873
DB 642 DYIREHKNIGSQYLLNMCVQIAGKMNVLERHVMHVRDLAARNVLVKTIPQHVKITDFGLA 701
QY 874 RLLDIDETEHADGKVPDKWMALESILRRRTHQSDVMSYGVTVWELMTGAKPYDGIP 933
DB 702 KOLGADEKEYHAEGKVPDKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIP 761
QY 934 AREIPDLLEKGERLPQPPICITDVTVMVCKMIDSECRPRRELVSFESRNPARDPQFV 993
DB 762 ASEISSVLEKGERLPQPPICITDVTVMVCKMWSGADSRPKPRELIAEFSKRNARDPPRYL 821

QY 994 VIQ-NEDLGPASPLDSTFYRSLLEDMDGLVDABEYLVPOQGFPCDPAPGAGMVHHR 1052
DB 822 VIQGDREHMLPSFTDTSKFTLMEEDMEDIVDAEYLVPHQGF-----866
QY 1053 HRSSSTRSGGDLTLGLEPSEBEAPRSPLAPSEGAGSDVFDGLGMAAKGLQSLPHOP 1112
DB 867 -NSPST-----SRTLPLSSLSATSN-----NSATKCIDRNGGH-- 898
QY 1113 SPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEVYNQPDVVRPOPSPREGPLPAARPAGAT 1172
DB 899 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYNYISLT 936
QY 1173 -LERAKTLSPGKNGVVKVFAFGGAVENPEYL 1203
DB 937 AISKLPMDSRYN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scolding P., Vennetrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CAA30024.1; -
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCE8CCA0F8AF4 CRC64;

Query Match 24.0%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 1.2e-115;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 584 GEADQCVACAHYKDPFCVAPCPGSKVPLSYMPIWKFPDESGACQPCPINCTHSCVDL 643
DB 1 GP--DHCKMCAHFIDGPHCVKACPAVLGENDTL-VKDYADANAVCOLCHPNCCTCKG 57
QY 644 DDKGPAPQORASPLTSIVSAVV-GILLVVLGVVFGILIKRROOKIRKVTMRRLLOETEL 702
DB 58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVGIGLYLRRR-HIVRKRTLRLQLQEREL 113
QY 703 VEPLTPSGAMPNQAQRILKETELRKVKVLGSGAFGTGVYKGIWIPGENVKIPVALIKVL 762
DB 114 VEPLTPSGEAPNQAHLRIKETEFKKVKVLGFGAFGTGVYKGLWIPEGEKVTIPVALIKEL 173
QY 763 ENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRRGR 822
DB 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLVTOLMPYGCLLDIYREHKN 233
QY 823 LGSQDILLNMCQIAGKMSYLEDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETE 882
DB 234 LGSQYLLNMCVQIAGKMNVLERHVLVHRDLAARNVLVKTIPQDVKITDFGLAKQLGADEKE 293

```
QY 883 YHAGGKVPKMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLE 942
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 294 YHAGGKVPKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISVLE 353
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 943 KGERLPQPPCTTIDVTMIMVKMWIDSECRPRRELVSFSPRWARDQPRVVIQ-NEDLG 1001
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 354 KGERLPQPPCTTIDVTMIMVKMWSDADSRPKRELIAEFKQARDPPRYLVITQGDHMH 413
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1002 PASPLDSTFVRSILLEDDMGDLVDAEYLVPQGGFFCPDPAPCAGGWHVHRHSSSTRSG 1061
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 414 LPSPTDSKFRVTEWEEDEMDIVDAEYLVPHQGF-----NSPST--- 454
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1062 GGDLTGLLEPSEEEAPRSP-----APSEGAGSDVFDGDLGMAAKGLQSLPHDRPSPLQ 1116
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 455 -----SRTPLLSLSATSNNNSATNCIDRNG-----H----- 481
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1117 RYSEDTVPLPSETDGVAPLTCSPQPEYVQNDVPRQPPSPREGPLPAARPAAGAT-LER 1175
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKPBSTAMVQIQIYISLTAISK 523
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1176 AKTLSPCKNGVVKDVFAGGAVENPEYL 1203
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 524 LPWDSRYQN-----SHSTAVDNPEYL 544
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

RESULT 15

```
Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TReMBUrel. 12, Created)
DT 01-NOV-1999 (TReMBUrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBUrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7TAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearfall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiri L.M., Staublei F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barch G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
```

```
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;
Query Match 22.3%; Score 1530.5; DB 11; Length 655;
Best Local Similarity 44.0%; Pred. No. 7.2e-107;
Matches 284; Conservative 100; Mismatches 232; Indels 29; Gaps 7;
QY 11 LLLALLPPGAA--STQVCTGTDMKRLPASPETHLDMLRHLYGCGVQVGNLEUYLPTN 68
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 LLTALCAAGALEEKVCGQTSNRLTQLGTFEDHFLSLQRMNVCVVLGNLEITYVQRN 73
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 69 ASLSFLDIOEVOGYVLIHNOVQVPLQRLRIVRGTLQFEDNYALAVLNDGDLNNTTP 128
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 YDLSFLTKTQEVAGYVLIHNTVERPLENLQIRGNALYENTYALAILSN----- 124
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 129 VTGASPGGLRELQLRLSLTEILKGVLIQRNPLCYODTILKWDNFNFTVFWLVRPKVSA 188
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 -YGNRTGLRELPMRLQELIGAVRFSNNPILCNMDTIQWRDI-----VQNVFM 173
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 189 SH-----LENRSRACHPCSPCKMGSCWGESSEDCQLTRTVCAAGCA-RCKGPLPTDCCH 243
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 174 SNMSMDLQSHSSCPKCDPSCPNWSCWGGGEENCQKLTIIAQCSHRGRSPSCCH 233
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 244 EQCAAGCTGPKGSDCLACILHFNHSGICELHCPALVTVNTDTTFESMPNPEGRYTFGASCVT 303
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 234 NQCAAGCTGPRSDCLVCQKFQDEATCKDTCPPLMLYNPTTYQMDVNPPEGKYSFGATCVK 293
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 304 ACPYNYLSTDVGSCTLVCPHLHNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVAVT 363
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 294 KCPRTVTVTHGSCVTRACGPDYIEV-EEDGIRKCKKDCPCRVCNIGIGFEDKDTUSIN 352
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 364 SANIQFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQVLFETLEEITGYLYISAWPD 423
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 353 ATNKHFKYCTAISGDLHLIPVAFKGSFTRTPPLDPRELEILKTVKEITGFLLIQAWPD 412
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 424 SLPDLSVFQNLQVIRGIRLHNGAYSLSLQGLGTSWLGRLSLRGLSGGLAIHHTHLCFV 483
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 413 NMTDLHAFENLEIRGTRTKHQHGFSLAVVGLNITSLGRLSLKEISDGVITSGNRNLCYA 472
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 484 HTPVPDQLFRNPHOALLHTANRDEDCVGBGLACHOLCARGHCWGCPGTPCVCNCSQFLRG 543
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 473 NTINWKLFGTPNQKTQKIMNRAEKCKKAVNHVNCPLCSSEGCGPPEPRDCVQNSVRG 532
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 544 QECVEECRLVQLPREYVFNARHCLPCHPECPQNGSVTCFPGPADQCACAHYKDPFV 603
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 533 RECCEKCNILEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCV 592
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 604 ARCPGSGVKDLSYMTWKVPDEEGACQPCPINTHSCVDLDDKGC 648
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 593 KTCPCAGIMGNNTL-VMKYADANNVCHLCHANCTYGCAGPLQGC 636
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```


Search completed: July 22, 2003, 09:00:33
Job time : 53.5887 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-149-163-14

Perfect score: 6848

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 101002.*

```

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|--------------------|
| 1 | 6715 | 98.1 | 1255 | 21 | Human heregulin 2 |
| 2 | 6715 | 98.1 | 1255 | 22 | Human tyrosine kin |
| 3 | 6715 | 98.1 | 1255 | 22 | HER2 transgene pla |
| 4 | 6715 | 98.1 | 1255 | 23 | Human HER2 (ErbB2) |
| 5 | 6709 | 98.0 | 1255 | 17 | HER-2/neu protein. |
| 6 | 6709 | 98.0 | 1255 | 20 | Human HER-2/neu on |
| 7 | 6709 | 98.0 | 1255 | 21 | Human HER-2/neu pr |
| 8 | 6709 | 98.0 | 1255 | 21 | Amino acid sequenc |
| 9 | 6709 | 98.0 | 1255 | 22 | Human HER-2/neu pr |
| 10 | 6709 | 98.0 | 1255 | 22 | HER2/neu amino aci |

| | | | | | | |
|----|--------|------|------|----|----------|--------------------|
| 11 | 6709 | 98.0 | 1255 | 23 | AAE24067 | Human Her-2 protei |
| 12 | 6709 | 98.0 | 1255 | 23 | AAE20479 | Human Her-2/neu pr |
| 13 | 6709 | 98.0 | 1255 | 23 | AAE20479 | Human Her-2/neu on |
| 14 | 6709 | 98.0 | 1255 | 23 | AAU77114 | Human Her-2/neu po |
| 15 | 6666 | 97.3 | 1433 | 14 | AAE39568 | Sequence of c-erbB |
| 16 | 6545 | 95.6 | 1433 | 14 | AAU98923 | Human breast canc |
| 17 | 6392 | 93.3 | 1200 | 21 | AAE21208 | Human HER-2/neu pr |
| 18 | 5928.5 | 86.6 | 1256 | 21 | AAE21199 | Rat Her-2/neu prot |
| 19 | 5928.5 | 86.6 | 1256 | 21 | AAE21199 | Rat Her-2/neu onco |
| 20 | 5912.5 | 86.3 | 1256 | 21 | AAE21206 | Mouse Her-2/neu pr |
| 21 | 5912.5 | 86.3 | 1256 | 22 | AAE21206 | Mouse Her-2/neu pr |
| 22 | 5912.5 | 86.3 | 1256 | 23 | AAE21206 | Amino acid sequenc |
| 23 | 4795 | 70.0 | 919 | 21 | AAE21203 | Human HER-2/neu on |
| 24 | 4795 | 70.0 | 919 | 23 | AAE21203 | Human HER-2/neu fu |
| 25 | 4062.5 | 59.3 | 920 | 23 | AAE21152 | Her-2/neu extracel |
| 26 | 4062.5 | 59.3 | 926 | 23 | AAE21152 | Mouse Her-2/neu ex |
| 27 | 3679 | 53.7 | 712 | 21 | AAE21204 | Human HER-2/neu ex |
| 28 | 3679 | 53.7 | 712 | 23 | AAE21204 | Human HER-2/neu fu |
| 29 | 3533 | 51.6 | 782 | 18 | AAE19764 | Her-2/neu extracel |
| 30 | 3531 | 51.6 | 653 | 21 | AAE21200 | Extracellular HER- |
| 31 | 3531 | 51.6 | 653 | 23 | AAE21200 | Human ErbB2 oncopr |
| 32 | 3493 | 51.0 | 645 | 22 | AAE20408 | Human ErbB2 extrac |
| 33 | 3493 | 51.0 | 645 | 22 | AAE20408 | Human ErbB2 extrac |
| 34 | 3428 | 50.1 | 951 | 21 | AAE44993 | DC8ecFv-erbB2EC fu |
| 35 | 3325 | 48.6 | 624 | 11 | AAE08222 | Extracellular port |
| 36 | 3164 | 46.2 | 1210 | 21 | AAE19259 | Amino acid sequenc |
| 37 | 3164 | 46.2 | 1210 | 21 | AAE19259 | Human EGF receptor |
| 38 | 3164 | 46.2 | 1210 | 23 | AAE23019 | Human Her-1 protei |
| 39 | 3164 | 46.2 | 1210 | 23 | AAE50768 | Human epidermal gr |
| 40 | 3162 | 45.2 | 1210 | 22 | AAE68420 | Amino acid sequenc |
| 41 | 3123 | 45.6 | 1210 | 23 | AAE51768 | Human epidermal gr |
| 42 | 3084 | 45.0 | 583 | 23 | AAE20483 | Human protein for |
| 43 | 3084 | 45.0 | 587 | 23 | AAE20481 | Human protein for |
| 44 | 3083 | 45.0 | 589 | 23 | AAE20484 | Human protein for |
| 45 | 3083 | 45.0 | 600 | 23 | AAE20482 | Human protein for |

ALIGNMENTS

RESULT 1

AAE20482
ID AAE20482 standard; Protein; 1255 AA.

XX AAE20482;

AC AAE20482;

XX 10-AUG-2000 (first entry)

DT 10-AUG-2000 (first entry)

XX Human heregulin 2 (Her2).

DE Human heregulin 2 (Her2).

XX Human heregulin 2 (Her2).

KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;

KW self-protein; cancer; breast cancer; prostate cancer;

KW cell-associated peptide antigen; foreign epitope.

XX Homo sapiens.

OS Homo sapiens.

XX Key

FH Domain

FT Location/Qualifiers

FT 1..173

FT /label= N-terminal

FT /note= "mature polypeptide"

FT 5..25

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT 59..73

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT 103..117

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT 149..163

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT 174..323

FT Domain

QY 777 AYVAVGSPVVSRLGICLTSTVLTQMPYGCCLLDHVRNRRGLSQDLLNWCMA 836
DB 771 AYVAVGSPVVSRLGICLTSTVLTQMPYGCCLLDHVRNRRGLSQDLLNWCMA 830
QY 837 KGMSVLEDRVLRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKKVPIKMA 896
DB 831 KGMSVLEDRVLRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKKVPIKMA 890
QY 897 LESILRRRFTHQSDVWSYGVTVWELMTFCAPYDGIPIAREIPDLLEKGERLPPOPICTID 956
DB 891 LESILRRRFTHQSDVWSYGVTVWELMTFCAPYDGIPIAREIPDLLEKGERLPPOPICTID 950
QY 957 YVIMVCKWIMIDSECRPRELVSFNSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1016
DB 951 YVIMVCKWIMIDSECRPRELVSFNSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1010
QY 1017 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGVMVHRHSSSTRSGGDLTLGLEPSEBEA 1076
DB 1011 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGVMVHRHSSSTRSGGDLTLGLEPSEBEA 1070
QY 1077 PRSPLAPSEGAGSDVFDGDLGMAAKGLQSLPHTHDPSPLOQRYSEDPVPLPSETDGYVAP 1136
DB 1071 PRSPLAPSEGAGSDVFDGDLGMAAKGLQSLPHTHDPSPLOQRYSEDPVPLPSETDGYVAP 1130
QY 1137 LTCSPQPEYVNPQVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVPAFGGA 1196
DB 1131 LTCSPQPEYVNPQVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVPAFGGA 1190
QY 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
DB 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
QY 1257 LDVPV 1261
DB 1251 LDVPV 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX AC AAE12130;
XX DT 18-DEC-2001 (first entry)
XX DE Human tyrosine kinase-type receptor, HER-2.
XX KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Region 774..782
XX FT /note= "Antigenic epitope"
XX PN WO200168677-A2.
XX PD 20-SEP-2001.
XX XX 16-MAR-2001; 2001WO-US40328.
XX PF 16-MAR-2000; 2000US-0527487.
XX PR (GENZ) GENZYME CORP.
XX PA Nicolette CA;
XX PI
XX DR WPI; 2001-616284/71.
XX DR N-PSDB; AAD19731.
XX XX

PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT .
XX Claim 4; Page 63-67; 69pp; English.
PS The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX Sequence 1255 AA;
SQ Query Match 98.1%; Score 6715; DB 22; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1242; Conservative 6; Mismatches 3; Indels 14; Gaps 3;
QY 1 MELAACRWGLLALLPFGAASQTCTGTDMLRLPASPETHLDMRLHYQGCQVQGNL 60
DB 1 MELAACRWGLLALLPFGAASQTCTGTDMLRLPASPETHLDMRLHYQGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVRQVPLQRLRIVRGTQFEDNALAVDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVRQVPLQRLRIVRGTQFEDNALAVDNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCQVQDTILWKDF---NNFTV 177
DB 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCQVQDTILWKDF---NNFTV 177
QY 178 SFWLRVPKVSASHLE-NRSRACHPCSPMKSGRCWGESSEDCQSLTRTYCAGGCARCKGP 236
DB 178 -----QLALTLDITNRSRACHPCSPMKSGRCWGESSEDCQSLTRTYCAGGCARCKGP 230
QY 237 LPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESHPNPGRYT 296
DB 231 LPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESHPNPGRYT 290
QY 297 FGASCVTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHL 356
DB 291 FGASCVTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHL 350
QY 357 REVRAVTSANTIOEFAGCKKIFGSLAFLPESDGPASNTAPLOEQLOVFTLEEITGYL 416
DB 351 REVRAVTSANTIOEFAGCKKIFGSLAFLPESDGPASNTAPLOEQLOVFTLEEITGYL 410
QY 417 YISAWPDSLPLDSVFNQVIRGRILLHNGAYSLTLQGLGISWLGRLSRLSGSLALIIHH 476
DB 411 YISAWPDSLPLDSVFNQVIRGRILLHNGAYSLTLQGLGISWLGRLSRLSGSLALIIHH 470
QY 477 NTHLCFVHTVPWDOLFRNPHOALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTOCVN 536
DB 471 NTHLCFVHTVPWDOLFRNPHOALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTOCVN 530
QY 537 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCACAHY 596
DB 531 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCACAHY 590
QY 597 KDPFFCVARCPGSKVPLDSVNPYIWKFPDEGACQPCPINCTHSCVDLDDKGPACQORASP 656
DB 591 KDPFFCVARCPGSKVPLDSVNPYIWKFPDEGACQPCPINCTHSCVDLDDKGPACQORASP 650

```
QY 657 LTSIVSAVVGILLVVLGVVFGILIKRROOKIRKYIMRRLLQETELVEPLTPSGAMPNOA 716
DB 651 LTSIVSAVVGILLVVLGVVFGILIKRROOKIRKYIMRRLLQETELVEPLTPSGAMPNOA 710
QY 717 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKBILDE 776
DB 711 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKBILDE 770
QY 777 AYVMAGVGSPPYVSRLLGICLTSTVOLVTOLMPYGCILLDHVRENRRGLSGODLLNWCMOIA 836
DB 771 AYVMAGVGSPPYVSRLLGICLTSTVOLVTOLMPYGCILLDHVRENRRGLSGODLLNWCMOIA 830
QY 837 KGMYSILEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWWA 896
DB 831 KGMYSILEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWWA 890
QY 897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTID 956
DB 891 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTID 950
QY 957 VYIMVUKWIMIDSECRPRELVESEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYBSLLE 1016
DB 951 VYIMVUKWIMIDSECRPRELVESEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYBSLLE 1010
QY 1017 DDDMGDLVDAEYLVPOQGFCCPDPAAGGVMVHHRSSSTRSGGGDLTLGLEPSEEEA 1076
DB 1011 DDDMGDLVDAEYLVPOQGFCCPDPAAGGVMVHHRSSSTRSGGGDLTLGLEPSEEEA 1070
QY 1077 PRSPLAPSEGAGSDVFDGLMGAAKGLQLPHTDPSPLQRYSEDPTVPLPSSTDGVYAP 1136
DB 1071 PRSPLAPSEGAGSDVFDGLMGAAKGLQLPHTDPSPLQRYSEDPTVPLPSSTDGVYAP 1130
QY 1137 LTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTLPKNGVVKDVFAPFGGA 1196
DB 1131 LTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTLPKNGVVKDVFAPFGGA 1190
QY 1197 VENPEYLTPOGGAAPHPHPPAPSPAFDNLVYWDQDPPPERGAPPSTFKGTPTTAENPEYLG 1256
DB 1191 VENPEYLTPOGGAAPHPHPPAPSPAFDNLVYWDQDPPPERGAPPSTFKGTPTTAENPEYLG 1250
QY 1257 LDVPV 1261
DB 1251 LDVPV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC
XX
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX WO200100244-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-US17229.
XX
XX 25-JUN-1999; 99US-0141316.
XX 16-MAR-2000; 2000US-0189844.
XX
XX (GETH ) GENENTECH INC.
XX
XX Erickson S, Schwall R;
PI
```

```
XX WPI: 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid.
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 98.1%; Score 6715; DB 22; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1242; Conservative 6; Mismatches 3; Indels 14; Gaps 3;
QY 1 MELAALCRNGLLIALLPGAASQTCTGDMKRLPASPEHLDMLRHLHYQGCVVQGNL 60
DB 1 MELAALCRNGLLIALLPGAASQTCTGDMKRLPASPEHLDMLRHLHYQGCVVQGNL 60
QY 61 EUTYLPNLSLFLQDIEVOGVYLIARNOVROVPLORLIRVGTQLFEDNYALAVLDNG 120
DB 61 EUTYLPNLSLFLQDIEVOGVYLIARNOVROVPLORLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQLRLSLTEILKGGVLIORNPOLCYQDTILWKDF---NNFTV 177
DB 121 DPLNNTPTVTGASPGGLRELQLRLSLTEILKGGVLIORNPOLCYQDTILWKDF---NNFTV 177
QY 178 SFMLRPVKVSASHLE-NRSRACHPCSPCMKSGRCSWSESSEDCOSLTRTVCAGCARCKGP 236
DB 178 -----QLALTIDTNRSRACHPCSPCMKSGRCSWSESSEDCOSLTRTVCAGCARCKGP 230
QY 237 LPTDCCHEOCAGCTGPKHSDCLACLFHNSHSGICELHCPALVTYNTDTTFESMPNPEGRYT 296
DB 231 LPTDCCHEOCAGCTGPKHSDCLACLFHNSHSGICELHCPALVTYNTDTTFESMPNPEGRYT 290
QY 297 FGASCVTACPNYVNSTDVGSCTLVCPHMQEVTAEQDTCRCEKSPCARVCYGLGMEHL 356
DB 291 FGASCVTACPNYVNSTDVGSCTLVCPHMQEVTAEQDTCRCEKSPCARVCYGLGMEHL 350
QY 357 REVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 416
DB 351 REVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 410
QY 417 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGRLSLRELGSGLALIH 476
DB 411 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGRLSLRELGSGLALIH 470
QY 477 NTHLCFVHTVPWDLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHGWGPGPTOCVN 536
DB 471 NTHLCFVHTVPWDLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHGWGPGPTOCVN 530
QY 537 CSQFLRGQCEVEECRVLQGLPREYVNAHCLPCHPECCQPNQSVTCFGEADQCVACAHY 596
DB 531 CSQFLRGQCEVEECRVLQGLPREYVNAHCLPCHPECCQPNQSVTCFGEADQCVACAHY 590
QY 597 KDPFFCVARCPGVKPDLSYMPIWKFPPDEGACQPCPINCTHSCVDLDDKGCBAEQRAS 656
DB 591 KDPFFCVARCPGVKPDLSYMPIWKFPPDEGACQPCPINCTHSCVDLDDKGCBAEQRAS 650
QY 657 LTSIVSAVVGILLVVLGVVFGILIKRROOKIRKYIMRRLLQETELVEPLTPSGAMPNOA 716
DB 651 LTSIVSAVVGILLVVLGVVFGILIKRROOKIRKYIMRRLLQETELVEPLTPSGAMPNOA 710
QY 717 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKBILDE 776
DB 711 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKBILDE 770
```

```
Db      711 QMRILKTELKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKELDE 770
QY      777 AYVAGVGSPPYVSRLIGICLTSTVOLVQLMPYGCLLDHVRENRLGSLDNLNWCQIA 836
Db      771 AYVAGVGSPPYVSRLIGICLTSTVOLVQLMPYGCLLDHVRENRLGSLDNLNWCQIA 830
QY      837 KMSYLEVDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKWWA 896
Db      831 KMSYLEVDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKWWA 890
QY      897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTID 956
Db      891 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTID 950
QY      957 VYIMVYKWMIDSECPRELVSESRMARDPQRFVWVTONEDLGPASPLDSTFYRSLLE 1016
Db      951 VYIMVYKWMIDSECPRELVSESRMARDPQRFVWVTONEDLGPASPLDSTFYRSLLE 1010
QY      1017 DDDMGDLVDAEYLVPOQGFCCPDPAAGGVMVHRRSSSTRSGGDLTLGLEPSEEEA 1076
Db      1011 DDDMGDLVDAEYLVPOQGFCCPDPAAGGVMVHRRSSSTRSGGDLTLGLEPSEEEA 1070
QY      1077 PRSPLAPSGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1136
Db      1071 PRSPLAPSGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1130
QY      1137 LTCSPQEVYNQPDVTPQPSREGPLPAARPAATLERAKTLSPGKNGVVKDVPFPGA 1196
Db      1131 LTCSPQEVYNQPDVTPQPSREGPLPAARPAATLERAKTLSPGKNGVVKDVPFPGA 1190
QY      1197 VENPEVLTTPQGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLG 1256
Db      1191 VENPEVLTTPQGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLG 1250
QY      1257 LDVPV 1261
Db      1251 LDVPV 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoeic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWOWSKI M.
XX
```

```
PI Erickson S, Schwall R, Sliwowski M;
XX
DR WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoeic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
Query Match 98.1%; Score 6715; DB 23; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1242; Conservative 6; Mismatches 3; Indels 14; Gaps 3;
QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGTMKRLPASPETHLDMLRHLHYGQCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGTMKRLPASPETHLDMLRHLHYGQCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIEVQGVLIHNVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEVQGVLIHNVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDITLWKDF---NNFTV 177
Db 121 DPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDITLWKDFHFNKND 177
QY 178 SFMLRVKPVYSASHLE-NRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAGGCARCKGP 236
Db 178 -----QLALTILDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAGGCARCKGP 230
QY 237 LPTDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
QY 297 FGASCVTACPNYVLTSDVGSCTLVCPHNVQVTAEDGTORCEKCKSPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYVLTSDVGSCTLVCPHNVQVTAEDGTORCEKCKSPCARVCYGLGMEHL 350
QY 357 REVRAVTSANIQEFAGCKKIFGSLAPFSPFSGDPAANTAPLOPEQLQVFETLEEITGYL 416
Db 351 REVRAVTSANIQEFAGCKKIFGSLAPFSPFSGDPAANTAPLOPEQLQVFETLEEITGYL 410
QY 417 YISAWPDSLPLDLSVFQNLQVIRGRILHNGAYSLITLOGLGISWLGRLSRELGLALIIH 476
Db 411 YISAWPDSLPLDLSVFQNLQVIRGRILHNGAYSLITLOGLGISWLGRLSRELGLALIIH 470
QY 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECYGEGLACHOLCARGHCWGPGTCQVN 536
Db 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECYGEGLACHOLCARGHCWGPGTCQVN 530
QY 537 CSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHY 596
Db 531 CSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHY 590
QY 597 KDPPECVACRCPGSKVPKDLISYMPFWKPEBEGACQPCPINCTHSCVDLDDKGCPEQRASP 656
```

| | | | |
|----------|--|---|------|
| Db | 591 | KDPFFCVCARCPGKPDLSYMPIWKPDDEGACQPCPINCTHSCVDLDDKGCRAEQRAS | 650 |
| Qy | 657 | LTSIVSAVVGILLVWLVGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQA | 716 |
| Db | 651 | LTSIVSAVVGILLVWLVGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQA | 710 |
| Qy | 717 | QMRILKETELURKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE | 776 |
| Db | 711 | QMRILKETELURKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE | 770 |
| Qy | 777 | AYVMAGVGSYPVSRLLGICLTSTVQLVTQMLPMPYGCCLLDHVRENRGRIGSODLLNWCNQIA | 836 |
| Db | 771 | AYVMAGVGSYPVSRLLGICLTSTVQLVTQMLPMPYGCCLLDHVRENRGRIGSODLLNWCNQIA | 830 |
| Qy | 837 | KGMSYLEDVLRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMA | 896 |
| Db | 831 | KGMSYLEDVLRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMA | 890 |
| Qy | 897 | LESILRRRFTHQSDVMSYGVTVWELMTFGAKPDYDGIIPAREIPDLEKGERLPPOPICTID | 956 |
| Db | 891 | LESILRRRFTHQSDVMSYGVTVWELMTFGAKPDYDGIIPAREIPDLEKGERLPPOPICTID | 950 |
| Qy | 957 | VYIMVWKWMDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE | 1016 |
| Db | 951 | VYIMVWKWMDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE | 1010 |
| Qy | 1017 | DDMGDLVDAEYLVPOQGFCDPAPGAGVWHRHRSSTRSGGGDLTLGLEPSBEEA | 1076 |
| Db | 1011 | DDMGDLVDAEYLVPOQGFCDPAPGAGVWHRHRSSTRSGGGDLTLGLEPSBEEA | 1070 |
| Qy | 1077 | PRSPAPSEGAGSDVDFCDLGMGAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAP | 1136 |
| Db | 1071 | PRSPAPSEGAGSDVDFCDLGMGAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAP | 1130 |
| Qy | 1137 | LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGA | 1196 |
| Db | 1131 | LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGA | 1190 |
| Qy | 1197 | VENPEYLTPOGGAAPHPHPPAPSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPYLG | 1256 |
| Db | 1191 | VENPEYLTPOGGAAPHPHPPAPSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPYLG | 1250 |
| Qy | 1257 | LDVPV 1261 | |
| Db | 1251 | LDVPV 1255 | |
| RESULT 5 | | | |
| ID | AAW01111 standard; Protein; 1255 AA. | | |
| XX | | | |
| AC | AAW01111; | | |
| XX | | | |
| DT | 01-JAN-1997 (first entry) | | |
| XX | | | |
| DE | HER-2/neu protein. | | |
| XX | | | |
| KW | HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase; | | |
| KW | breast cancer; ovary cancer; colon cancer; lung cancer; | | |
| KW | prostate cancer; immunisation; tumour; vaccine; vector. | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |
| PH | Key Location/Qualifiers | | |
| FT | Domain 676..1255 | | |
| FT | /label= Intracellular domain | | |
| XX | /note= "claimed domain, useful for immunisation" | | |
| PN | WO9630514-A1. | | |
| PD | 03-OCT-1996. | | |
| XX | | | |

| | | | |
|----|---|--|---------------------------------|
| PF | 28-MAR-1996; | 96WO-US01689. | |
| XX | | | |
| PR | 31-MAR-1995; | 95US-0414417. | |
| XX | | | |
| PA | (UNIW) UNIV WASHINGTON. | | |
| XX | | | |
| PI | Cheever MA, Disis ML; | | |
| XX | | | |
| DR | WPI; 1996-455361/45. | | |
| DR | N-PSDB; AAT40739. | | |
| XX | | | |
| PT | DNA encoding HER-2-neu poly:peptide(s) - used for prevention or | | |
| PT | treatment of malignancies with which the HER-2/neu oncogene is | | |
| PT | associated | | |
| XX | | | |
| XX | Claim 2; Page 56-61; 71pp; English. | | |
| XX | | | |
| CC | Human HER-2/neu protein (AAM01111), also called p185 or c-erbB2, is | | |
| CC | the product of the HER-2/neu oncogene (see also AAT40739). The | | |
| CC | protein is over-expressed in various cancers, including breast, | | |
| CC | ovarian, colon, lung and prostate. The intracellular domain of the | | |
| CC | protein can be used to immunise an animal against a malignancy with | | |
| CC | which the oncogene is associated. The polypeptide can be produced | | |
| CC | in transformed host cells for use in immunisation. Alternatively, | | |
| CC | animal cells are transfected in vivo or ex vivo with a viral vector | | |
| CC | that directs expression of the polypeptide. | | |
| XX | | | |
| SQ | Sequence 1255 AA; | | |
| | Query Match | 98.0%; | Score 6709; DB 17; Length 1255; |
| | Best Local Similarity | 98.0%; | Pred. No. 0; |
| | Matches 1240; Conservative | 7; Mismatches | 4; Indels 14; Gaps 3; |
| Qy | 1 | MELAAALCRWGLLLALLPPGAASQVCTGTDKMLRSPASPTHLDMLRHLVGGCQVQGNL | 60 |
| Db | 1 | MELAAALCRWGLLLALLPPGAASQVCTGTDKMLRSPASPTHLDMLRHLVGGCQVQGNL | 60 |
| Qy | 61 | ELTYLPTNASLSFLQDIEQVGVVLIAHNOVQVPLQRLRIVRGTQLFEDNYALAVLDNG | 120 |
| Db | 61 | ELTYLPTNASLSFLQDIEQVGVVLIAHNOVQVPLQRLRIVRGTQLFEDNYALAVLDNG | 120 |
| Qy | 121 | DPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTTLWKDF---NNFTV | 177 |
| Db | 121 | DPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTTLWKDFIHKNN--- | 177 |
| Qy | 178 | SFWLRVPKVSASHLE-NRSRACHPCSPMKGSRGWGESSEDCQSLTRTVGAGGCARCKGP | 236 |
| Db | 178 | -----QLALTLDITNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVGAGGCARCKGP | 230 |
| Qy | 237 | LPTDCCHEOCAAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYT | 296 |
| Db | 231 | LPTDCCHEOCAAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYT | 290 |
| Qy | 297 | FGASCVTACPNYLLSTDVSGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHL | 356 |
| Db | 291 | FGASCVTACPNYLLSTDVSGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHL | 350 |
| Qy | 357 | REVRVTSANIQBFAGCKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEETIGYL | 416 |
| Db | 351 | REVRVTSANIQBFAGCKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEETIGYL | 410 |
| Qy | 417 | YISAWPDSLPLDSVFONLOVIRGRILHNGAYSILTQGLGISWLGRLSRELGSGLALIH | 476 |
| Db | 411 | YISAWPDSLPLDSVFONLOVIRGRILHNGAYSILTQGLGISWLGRLSRELGSGLALIH | 470 |
| Qy | 477 | NTHLCFVHTVPWDQLFRNPHQALLHTANRDECEVCGEGLACHQLCARGHGWGPGPTQCVN | 536 |
| Db | 471 | NTHLCFVHTVPWDQLFRNPHQALLHTANRDECEVCGEGLACHQLCARGHGWGPGPTQCVN | 530 |
| Qy | 537 | CSQFLRGQECVCECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCAV | 596 |
| Db | 531 | CSQFLRGQECVCECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCAV | 590 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|-----|--------|-----|---|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|---|---|-----|-----|---|-----|
| Qy | 597 | KDPFCV | ARC | B | SGV | K | P | D | L | S | Y | N | P | I | W | K | F | D | E | E | G | A | C | O | P | C | I | N | T | H | S | C | V | D | L | D | D | K | G | C | P | A | E | R | A | S | P | 656 | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 591 | KDPFCV | ARC | B | SGV | K | P | D | L | S | Y | N | P | I | W | K | F | D | E | E | G | A | C | O | P | C | I | N | T | H | S | C | V | D | L | D | D | K | G | C | P | A | E | R | A | S | P | 650 | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Qy | 657 | LTSIVS | AV | V | G | I | L | L | V | V | L | G | V | F | G | I | L | K | R | O | Q | I | R | K | Y | T | M | R | R | L | L | O | E | T | E | L | V | E | P | L | T | P | S | G | A | M | P | N | Q | A | 716 | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 651 | LTSII | SA | V | G | I | L | L | V | V | L | G | V | F | G | I | L | K | R | O | Q | I | R | K | Y | T | M | R | R | L | L | O | E | T | E | L | V | E | P | L | T | P | S | G | A | M | P | N | Q | A | 710 | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Qy | 717 | QMRIL | K | E | T | E | R | K | V | K | V | L | G | S | A | F | T | V | Y | K | G | I | W | P | D | G | E | N | V | K | I | P | V | A | K | V | L | R | E | N | T | S | P | K | A | N | K | E | L | D | E | 776 | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 711 | QMRIL | K | E | T | E | R | K | V | K | V | L | G | S | A | F | T | V | Y | K | G | I | W | P | D | G | E | N | V | K | I | P | V | A | K | V | L | R | E | N | T | S | P | K | A | N | K | E | L | D | E | 770 | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Qy | 777 | AYN | A | G | V | G | S | P | V | S | R | L | L | G | I | C | L | T | S | T | V | O | L | T | O | L | M | P | Y | C | L | L | D | H | V | R | E | N | R | G | R | L | S | O | D | L | L | N | C | M | O | I | A | 836 |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 771 | AYN | A | G | V | G | S | P | V | S | R | L | L | G | I | C | L | T | S | T | V | O | L | T | O | L | M | P | Y | C | L | L | D | H | V | R | E | N | R | G | R | L | S | O | D | L | L | N | C | M | O | I | A | 830 |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Qy | 837 | KGMS | L | E | D | V | R | L | V | H | R | D | L | A | A | R | N | V | L | V | K | S | P | N | H | V | K | I | T | D | F | L | A | R | L | I | D | I | D | E | T | E | H | A | D | G | K | V | P | I | K | M | | |

RESULT 6

RESULTS 8
AAW92406
ID AAW92406 standard; Protein: 1255 AA.

XX
AC AAW92406;

XX
DT 21-APR-1999 (first entry)

XX
DE Human HER-2/neu oncogene protein.

| | |
|----|---|
| XX | |
| KW | HER-2/neu; oncogene; immune response; T cell; B cell; immunisation; |
| KW | malignancy; treatment; tumour. |

XX
OS Homo sapiens.

| XX | Key | Location/Qualifiers |
|----|--------|---------------------|
| FH | Region | 676..1255 |
| FT | | |

FT /note= "region which elicits immune response"

XX
PN US5869445-A.

XX
PD 09-FEB-1999.

[illegible]

QY 597 KDPFFCVARCPGKVPDLSYMPITWKFDEEGACOPCPINCTHSCVDLDDKGCAPAEORASP 656
DB |||||
QY 591 KDPFFCVARCPGKVPDLSYMPITWKFDEEGACOPCPINCTHSCVDLDDKGCAPAEORASP 650
DB |||||
QY 657 LTSIVSAVGLLVVGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQA 716
DB |||||
QY 651 LTSIISAVGTLVVVGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQA 710
DB |||||
QY 717 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETLDE 776
DB |||||
QY 711 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETLDE 770
DB |||||
QY 777 AYVMAVGSPVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRGRIGSODLLNWCQIA 836
DB |||||
QY 771 AYVMAVGSPVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRGRIGSODLLNWCQIA 830
DB |||||
QY 837 KGSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDTEYHADGGKVPKMA 896
DB |||||
QY 831 KGSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDTEYHADGGKVPKMA 890
DB |||||
QY 897 LESILRRRFTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 956
DB |||||
QY 891 LESILRRRFTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 950
DB |||||
QY 957 VYIMVVKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1016
DB |||||
QY 951 VYIMVVKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1010
DB |||||
QY 1017 DDDMGDLVDAEYLVPOGFCPPDAPGAGGMVHRHRSSTRSGGDLTLGLSPSEEA 1076
DB |||||
QY 1011 DDDMGDLVDAEYLVPOGFCPPDAPGAGGMVHRHRSSTRSGGDLTLGLSPSEEA 1070
DB |||||
QY 1077 PRSLPASEGAGSDVFDGLGMGAAGKIQSLTHDPSPLOQYSDPTVPLPSETDGYVAP 1136
DB |||||
QY 1071 PRSLPASEGAGSDVFDGLGMGAAGKIQSLTHDPSPLOQYSDPTVPLPSETDGYVAP 1130
DB |||||
QY 1137 LTCSPQPEYVNPQDVRFPQPSREGPLPAARPAAGATLERAKTSLSPGKNGVVKDYFAFGA 1196
DB |||||
QY 1131 LTCSPQPEYVNPQDVRFPQPSREGPLPAARPAAGATLERAKTSLSPGKNGVVKDYFAFGA 1190
DB |||||
QY 1197 VENPEYLTPOGGAAPQHPHPPAFSPADNLYWDQDPPERGAPSTFKGTPTAENPEYL 1256
DB |||||
QY 1191 VENPEYLTPOGGAAPQHPHPPAFSPADNLYWDQDPPERGAPSTFKGTPTAENPEYL 1250
DB |||||
QY 1257 LDVPV 1261
DB |||||
QY 1251 LDVPV 1255
DB |||||

RESULT 7

AAB21198

ID AAB21198 standard; protein; 1255 AA.

XX AC AAB21198;

XX DT 12-JAN-2001 (first entry)

XX DE Human HER-2/new protein.

XX KW Human; HER-2/new; oncogene; tyrosine kinase; cytostatic; vaccine;

XX KW breast cancer; prostate cancer; ovarian cancer; lung cancer;

XX KW colon cancer.

XX OS Homo sapiens.

XX PN WO200044899-A1.

XX PD 03-AUG-2000.

XX PF 28-JAN-2000; 2000WO-US02164.

XX PR 29-JAN-1999; 99US-0117976.

XX XX

PA (CORI-) CORIXA CORP.
PA (SMIK) SMITHKLINE BEECHAM.
PI Cheever MA, Gheysen D;
XX WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX HER-2/new extracellular domain/phosphorylation domain fusion proteins
XX useful for vaccinating against breast, ovarian, colon, lung and
XX prostate cancers -
XX Claim 52; Fig 7; 128pp; English.
XX The present sequence is the human HER-2/new protein. It is a member of
XX the tyrosine kinase family of receptor-like glycoproteins and shows
XX homology to the epidermal growth factor receptor (EGFR). It probably
XX plays a part in cell growth and/or differentiation. The HER-2/new
XX gene is an oncogene. An HER-2/new fusion protein comprising a
XX HER-2/new extracellular domain fused to a HER-2/new phosphorylation
XX domain may be used to treat or prevent cancer by eliciting or
XX enhancing an immune response to the HER-2/new protein. It may be used
XX to treat malignancies such as breast, ovarian, colon, lung and
XX prostate cancers, and may be used as an antigen to vaccinate against
XX these neoplasias.
SQ Sequence 1255 AA;

Query Match 99.0%; Score 6709; DB 21; Length 1255;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;
QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDKMLRASPETHLDMLRHLGYCCVVOGNL 60
DB |||||
QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDKMLRASPETHLDMLRHLGYCCVVOGNL 60
DB |||||
QY 61 ELTYLPTNASLSFLQDIOEVQYVLIAHNOVQVPLQRLRIVRGTLQFEDNVALAVLNG 120
DB |||||
QY 61 ELTYLPTNASLSFLQDIOEVQYVLIAHNOVQVPLQRLRIVRGTLQFEDNVALAVLNG 120
DB |||||
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPQLCYQDTILWKDF---NNFTV 177
DB |||||
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPQLCYQDTILWKDF---NNFTV 177
DB |||||
QY 178 SFWLRVPKVSASHLE-NRSPACHPCSPCKGSRGCSGESSDCQSLTRTVCAAGCARCKGP 236
DB |||||
QY 178 SFWLRVPKVSASHLE-NRSPACHPCSPCKGSRGCSGESSDCQSLTRTVCAAGCARCKGP 230
DB |||||
QY 237 LPTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 296
DB |||||
QY 231 LPTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 290
DB |||||
QY 297 FGASCVTACPNYLSLTDVGSCTLVCPHNOEVTADGTQRCCKSKPCARVCYGLGMEHL 356
DB |||||
QY 291 FGASCVTACPNYLSLTDVGSCTLVCPHNOEVTADGTQRCCKSKPCARVCYGLGMEHL 350
DB |||||
QY 357 REYRAVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLOPELOVFTLEITGYL 416
DB |||||
QY 351 REYRAVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLOPELOVFTLEITGYL 410
DB |||||
QY 417 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSLTQGLGSLMGLRSRLRELSGLALH 476
DB |||||
QY 411 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSLTQGLGSLMGLRSRLRELSGLALH 470
DB |||||
QY 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCGPGPTQCVN 536
DB |||||
QY 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCGPGPTQCVN 530
DB |||||
QY 537 CSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHY 596
DB |||||
QY 531 CSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHY 590
DB |||||
QY 597 KDPFFCVARCPGKVPDLSYMPITWKFDEEGACOPCPINCTHSCVDLDDKGCAPAEORASP 656

Db 591 KDPFFCVCARCGSKVRLDLYMPFWKFPDEBEGACQPCINCTHSCVDLDDKGCFAEQRASP 650
Qy 657 LTSISAVVGIILLVVLGVVFGILIKRQOKIRKYTMRLLOTELVEPLTPSGAMPNOA 716
Db 651 LTSISAVVGIILLVVLGVVFGILIKRQOKIRKYTMRLLOTELVEPLTPSGAMPNOA 710
Qy 717 QMRILKETELRKVKVLGSGAFGTYYKGIWIPGENVKIPVAIKVLRNTPSPKANKBILDE 776
Db 711 QMRILKETELRKVKVLGSGAFGTYYKGIWIPGENVKIPVAIKVLRNTPSPKANKBILDE 770
Qy 777 AYVMAGVGSPPVSRLLGICLTSTVOLATVOTLMPYGCILLDHVRENRLGSGDILLNWCQJIA 836
Db 771 AYVMAGVGSPPVSRLLGICLTSTVOLATVOTLMPYGCILLDHVRENRLGSGDILLNWCQJIA 830
Qy 837 KGMYSLEEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDIDEYHADGGKVPKMA 896
Db 831 KGMYSLEEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDIDEYHADGGKVPKMA 890
Qy 897 LESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITD 956
Db 891 LESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITD 950
Qy 957 VYMIWVKWIMIDSECRPRELVSFERNARDPQRFVWIONEDLGPASPLDSTFYRSILLE 1016
Db 951 VYMIWVKWIMIDSECRPRELVSFERNARDPQRFVWIONEDLGPASPLDSTFYRSILLE 1010
Qy 1017 DDDMGDLVDAEYLVPOQGFCDPAPGAGGMVHHRSSSTRSGGDLTLGLPSEEEA 1076
Db 1011 DDDMGDLVDAEYLVPOQGFCDPAPGAGGMVHHRSSSTRSGGDLTLGLPSEEEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVWVDVAFGGA 1196
Db 1131 LTCSPQPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVWVDVAFGGA 1190
Qy 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAAPPSTFKTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAAPPSTFKTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 8

AA84780
ID AAY84780 standard; Protein; 1255 AA.

XX AC AAY84780;

XX DT 08-AUG-2000 (first entry)

XX DE Amino acid sequence of the SPLICE erbb-2 receptor protein.

XX KW SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.

XX OS Homo sapiens.

XX PN WO200020579-A1.

XX XX

XX PD 13-APR-2000.

XX PF 01-OCT-1999; 99WO-CA00912.

XX PR 02-OCT-1998; 98US-0165192.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PA

XX

PI

XX

DR

XX

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

(UYMC-) UNIV MCMASTER.

Muller WJ, Siegel PM;

WPI: 2000-303768/26.

N-PSDB; AAA14812.

Nucleic acid encoding an erbb 2 receptor protein designated SPLICE

erbb-2, inhibitors of the protein are useful for treatment of cancer

Claim 3; Fig 2; 60pp; English.

The present sequence represents a SPLICE erbb-2 receptor protein. The

protein has an in-frame deletion of 16 amino acids, 2 of which are

conserved cysteine residues, compared to the unspliced protein. The

erbb-2 polynucleotide is used to construct probes for detecting

disorders of cell transformation such as cancer. Antibodies to the

protein may be used to detect SPLICE erbb-2 in a sample. Agents

(e.g. antisense oligonucleotides) which inhibit the expression of

SPLICE erbb-2 are useful for reducing tumor cell proliferation and

treating cancer. Substances which stimulate SPLICE erbb-2 are useful

for treating conditions involving damaged cells including conditions

in which degeneration of tissue occurs, such as arthropathy, bone

resorption, inflammatory diseases, degenerative disorders of the

central nervous system and wound healing.

Sequence 1255 AA;

Query Match

Best Local Similarity 98.0%; Score 6709; DB 21; Length 1255;

Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;

Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPEHLDMLRHLYGGQVQVGNL 60

Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPEHLDMLRHLYGGQVQVGNL 60

Qy 61 EUTYLPNTASLSFLQDIQEVQVGLIAHQVQVPLQRLRIVRGTLFEDNYALVDNG 120

Db 61 EUTYLPNTASLSFLQDIQEVQVGLIAHQVQVPLQRLRIVRGTLFEDNYALVDNG 120

Qy 121 DFLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDF ---NNFTV 177

Db 121 DFLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDF --- 177

Qy 178 SFWLRVPKVSASHLE-NRSRACHPCSPMCKSGRCWGESSEDCOSLRTVTCAGGCARCKGP 236

Db 178 -----QUALTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVTCAGGCARCKGP 230

Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 296

Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 290

Qy 297 FGASCVTACPNYVLTSDVGSCTLVCPHNOEVTAEQTCRCKSPCARVCYGLGMEHL 356

Db 291 FGASCVTACPNYVLTSDVGSCTLVCPHNOEVTAEQTCRCKSPCARVCYGLGMEHL 350

Qy 357 REVRAVTSANIOEFAGCKKIFGSLAPLPSFDGDPASNTAPLOPELOVETLEEITGYL 416

Db 351 REVRAVTSANIOEFAGCKKIFGSLAPLPSFDGDPASNTAPLOPELOVETLEEITGYL 410

Qy 417 YISAWPDSLPLDLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIH 476

Db 411 YISAWPDSLPLDLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIH 470

Qy 477 NTHLCFVHTVPWDQLFRNPQALLHTANRPBECVCEGLACHQLCARGHCWGPGTQCVN 536

Db 471 NTHLCFVHTVPWDQLFRNPQALLHTANRPBECVCEGLACHQLCARGHCWGPGTQCVN 530

Qy 537 CSQFLRGQCEVEECVLOGLPREYVNAHCLCHPECQPNQSVTCFGEADQCVCAHY 596

Db 531 CSQFLRGQCEVEECVLOGLPREYVNAHCLCHPECQPNQSVTCFGEADQCVCAHY 590

Qy 597 KDPPFCVACRCPGSKVPSDLSYMPDIWKFPPDEGACQPCPINCTHSCVDLDDKGCFAEQRASP 656

| | | | |
|----|------|---|------|
| Db | 591 | KDPFFCVCPCSGVKPDLSTMPWKFPDEBGAQPCPINCTHSCVDLDDKGCFAQRASP | 650 |
| Qy | 657 | LTSIVSAVVGILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQA | 716 |
| Db | 651 | LTSIISAVVGILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQA | 710 |
| Qy | 717 | QMRILKETELURKVKVLGSGAFGYVYKGIWTPDGENVKIPVAIKVLRENTSPKANKETLDE | 776 |
| Db | 711 | QMRILKETELURKVKVLGSGAFGYVYKGIWTPDGENVKIPVAIKVLRENTSPKANKETLDE | 770 |
| Qy | 777 | AYVMAGVGSYPVSRLLGICLTSTVQLVTQLMPYGCLLDHYRENRGRGLSGODLLNWCMQIA | 836 |
| Db | 771 | AYVMAGVGSYPVSRLLGICLTSTVQLVTQLMPYGCLLDHYRENRGRGLSGODLLNWCMQIA | 830 |
| Qy | 837 | KGMSYLEDEVRLVHRDLAARNVLKVS PNHVKITDFGLARLLDIDETEVHADGGKVPIKNMA | 896 |
| Db | 831 | KGMSYLEDEVRLVHRDLAARNVLKVS PNHVKITDFGLARLLDIDETEVHADGGKVPIKNMA | 890 |
| Qy | 897 | LESILRRRFTHQSDVMSYGYVTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID | 956 |
| Db | 891 | LESILRRRFTHQSDVMSYGYVTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID | 950 |
| Qy | 957 | YYIMVWKCMWIDSECRPFRELSEFSEFMARDPQRFVVIQNEIDLGPASPLDSTFYFYSLLE | 1016 |
| Db | 951 | YYIMVWKCMWIDSECRPFRELSEFSEFMARDPQRFVVIQNEIDLGPASPLDSTFYFYSLLE | 1010 |
| Qy | 1017 | DDMGDILVDAEYILVPOQGFCCPDPAFCAGGVHHRHSSTRSGGGDLTLGLEPSEEEA | 1076 |
| Db | 1011 | DDMGDILVDAEYILVPOQGFCCPDPAFCAGGVHHRHSSTRSGGGDLTLGLEPSEEEA | 1070 |
| Qy | 1077 | PRSPAPSEGAGSDVFDGDLGMGAACKGLQSLPHTDPSPLQRYSEDPVPLPSETDGVVAP | 1136 |
| Db | 1071 | PRSPAPSEGAGSDVFDGDLGMGAACKGLQSLPHTDPSPLQRYSEDPVPLPSETDGVVAP | 1130 |
| Qy | 1137 | LTCSPQPEYVYNQPDVRPOPPSPREGPLUPAARPAGATLERAKTILSPGKNGVYVKDVFAGGA | 1196 |
| Db | 1131 | LTCSPQPEYVYNQPDVRPOPPSPREGPLUPAARPAGATLERPKTLSPGKNGVYVKDVFAGGA | 1190 |
| Qy | 1197 | VENPEXYLTPOGGAAPQPHPPAPFASPFDNLYYNDQDPPERGAPSTFKGTPTAENPEYLG | 1256 |
| Db | 1191 | VENPEXYLTPOGGAAPQPHPPAPFASPFDNLYYNDQDPPERGAPSTFKGTPTAENPEYLG | 1250 |
| Qy | 1257 | LDVPV 1261 | |
| Db | 1251 | LDVPV 1255 | |

RESULT 9
AAB85458
ID AAB85458 standard; Protein: 1255 AA:

| | |
|----|---|
| PI | Cheever MA, Hand-Zimmermann S; |
| XX | WPI; 2001-476112/51. |
| DR | N-PSDB; AAH23392. |
| XX | |
| XX | New antigen-presenting cells, useful as vaccines for eliciting or |
| PT | enhancing an immune response to HER-2/neu protein, particularly useful |
| PT | for treating or preventing cancer, e.g. breast cancer - |
| XX | |
| XX | Claim 2; Page 41-46; 49pp; English. |
| XX | |
| CC | The invention provides an isolated antigen-presenting cell, which |
| CC | expresses at least an immunogenic portion of a polypeptide that produces |
| CC | an immune response to HER-2/neu protein. The antigen-presenting cells are |
| CC | useful as vaccines for eliciting or enhancing an immune response to |
| CC | HER-2/neu protein, particularly in treating or preventing malignancies in |
| CC | which the HER-2/neu oncogene is associated. Specifically, these are |
| CC | useful for treating or preventing cancer, e.g. breast cancer, ovarian, |
| CC | colon, lung or prostate cancers. The present sequence represents |
| CC | the human HER-2/neu protein (also known as p185 or c-erbB2). |
| XX | |
| SQ | Sequence 1255 AA; |
| | |
| | Query Match 98.0%; Score 6709; DB 22; Length 1255; |
| | Best Local Similarity 98.0%; Pred. No. 0; |
| | Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3; |
| Qy | 1 MELAALCRWGLLALLPPGAASVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60 |
| Dd | |
| Dd | 1 MELAALCRWGLLALLPPGAASVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60 |
| Qy | 61 ELTYLPTNASLSFLQDIQEYGYVLIAHQVRQPQLRLIRVGTLQFEDNYALAVLDNG 120 |
| Dd | |
| Dd | 61 ELTYLPTNASLSFLQDIQEYGYVLIAHQVRQPQLRLIRVGTLQFEDNYALAVLDNG 120 |
| Qy | 121 DPLNNTTPTGASPGGLRELQLBSLTEILKGGVLIQRNPOLCYQDTRLHKDF---NNFTV 177 |
| Dd | |
| Dd | 121 DPLNNTTPTGASPGGLRELQLBSLTEILKGGVLIQRNPOLCYQDTRLHKDFPHKN---177 |
| Qy | 178 SFWLPRVPKSASHLE-NRGRACHPCSPMKGSRGWESSEDCOSLTRTVCGAGCARCKGP 236 |
| Dd | :::::- |
| Dd | 178 -----QLALTIDINRSPACHPCSPMKGSRGWESSEDCOSLTRTVCGAGCARCKGP 230 |
| Qy | 237 LPTDCCHEQCAACTGPKHSCLACLFHNHSGICEHLCPALVTYNTDTFESMPNEGRYT 296 |
| Dd | |
| Dd | 231 LPTDCCHEQCAACTGPKHSCLACLFHNHSGICEHLCPALVTYNTDTFESMPNEGRYT 290 |
| Qy | 297 FGASCVTACPYNLYSTDVGSCTLVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMEHL 356 |
| Dd | |
| Dd | 291 FGASCVTACPYNLYSTDVGSCTLVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMEHL 350 |
| Qy | 357 REVRVTSANIQEFAGCKKIFGSLAFLPESFDGDPASTAPLQPEOLQVFETLEEITGYL 416 |
| Dd | |
| Dd | 351 REVRVTSANIQEFAGCKKIFGSLAFLPESFDGDPASTAPLQPEOLQVFETLEEITGYL 410 |
| Qy | 417 YISAWPDSLPLDSVFQNLOVIRGRILLNHGAYSILTLOGLGISWLGSLRELSSGLALIHH 476 |
| Dd | |
| Dd | 411 YISAWPDSLPLDSVFQNLOVIRGRILLNHGAYSILTLOGLGISWLGSLRELSSGLALIHH 470 |
| Qy | 477 NTHLCFVHTVPMQDLFRNPHQALLHTANRPBEDEVGEGLACHQLCARGHCWGPPTQCYN 536 |
| Dd | |
| Dd | 471 NTHLCFVHTVPMQDLFRNPHQALLHTANRPBEDEVGEGLACHQLCARGHCWGPPTQCYN 530 |
| Qy | 537 CSOFLRGQECVEECRVLQGLPREYVNARHCLCHPECPQNGSVTCFGPEADQCACAHY 596 |
| Dd | |
| Dd | 531 CSOFLRGQECVEECRVLQGLPREYVNARHCLCHPECPQNGSVTCFGPEADQCACAHY 590 |
| Qy | 597 KDPFFCVARCPGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEORASP 656 |
| Dd | |
| Dd | 591 KDPFFCVARCPGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEORASP 650 |
| Qy | 657 LTSIVSAVGIILLVVGLVGFILIKRRQOKIRKYTMRELLQETELVEPLTPSGAMPNOA 716 |
| | |

Db 651 LTSIISAVVGVLLVVLVGVFGILIKRQKIRKYTRRLQLQETELVEPLTPSGAMPNOA 710
QY 717 QMRILKETELRKVKVLGSGAFGTGKIWIIPDGENVKIPVAIKVLRENTSPKANKEILDE 776
Db 711 QMRILKETELRKVKVLGSGAFGTGKIWIIPDGENVKIPVAIKVLRENTSPKANKEILDE 770
QY 777 AYVMAGVGSPPYVSRLLGICLTSTVQLVTQIMPGVCLLDHVRENRGLSGQDLLNWCQIA 836
Db 771 AYVMAGVGSPPYVSRLLGICLTSTVQLVTQIMPGVCLLDHVRENRGLSGQDLLNWCQIA 830
QY 837 KGMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMA 896
Db 831 KGMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMA 890
QY 897 LSSILARRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 956
Db 891 LSSILARRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950
QY 957 VYMIMVKWIMIDSECPRELVSERFMRDQRFVWIONEDLGPASPLDSTFYRSLE 1016
Db 951 VYMIMVKWIMIDSECPRELVSERFMRDQRFVWIONEDLGPASPLDSTFYRSLE 1010
QY 1017 DDDMGDLVDAEYLVLPQGGFCFDPAPGAGGMVHRRHSSTRSGGDLTLGLEPSEEA 1076
Db 1011 DDDMGDLVDAEYLVLPQGGFCFDPAPGAGGMVHRRHSSTRSGGDLTLGLEPSEEA 1070
QY 1077 PRSPLAPSGAGSDVPDGLGMAAKGLOSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSGAGSDVPDGLGMAAKGLOSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1130
QY 1137 LTCSPQPEVYNQDVRPQPPSPREGPLPAARPAAGTLERAKTILSPGKNGVKDVFAGGA 1196
Db 1131 LTCSPQPEVYNQDVRPQPPSPREGPLPAARPAAGTLERAKTILSPGKNGVKDVFAGGA 1190
QY 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
QY 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX
AC AAG88267;
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
FN WO200141787-A1.
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33591.
XX
PR 10-DEC-1999; 99US-0458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX Keogh E;
XX WPI; 2001-374995/39.

XX
PT
PT
XX
PS
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 1255 AA;

Query Match 98.0%; Score 6709; DB 22; Length 1255;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;
QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASBETHLDMLRHLYOGGVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASBETHLDMLRHLYOGGVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDF---NNFTV 177
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDF--- 177
QY 178 SFMLRVPKVSASHLE-NRSRACHPCSPCKGSRGWESSEDCOSLRTVTCAGGCARCKGP 236
Db 178 -----QUALTLIDNRSRACHPCSPCKGSRGWESSEDCOSLRTVTCAGGCARCKGP 230
QY 237 LPTDCCHQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCCHQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
QY 297 FGASCVTACPNYLTSTDVGSCTLVCPLNHNVQVTAEDGTORCEKSKPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYLTSTDVGSCTLVCPLNHNVQVTAEDGTORCEKSKPCARVCYGLGMEHL 350
QY 357 REVRAVTSANTQEFAGCKKIFGSLAPLPSFDGDPASNTAPLOPELOQVETLEEITGYL 416
Db 351 REVRAVTSANTQEFAGCKKIFGSLAPLPSFDGDPASNTAPLOPELOQVETLEEITGYL 410
QY 417 YISAMPDLSLPDLSVFNQLQVIRGRILHNGAYSLLTLQGLGISWGLRSLRELGSGLAIHH 476
Db 411 YISAMPDLSLPDLSVFNQLQVIRGRILHNGAYSLLTLQGLGISWGLRSLRELGSGLAIHH 470
QY 477 NTHLCFVHTVPMQDLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPGPTQCVN 536
|||||

Db 471 NTHLCFVHTVPMDLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVN 530
Qy 537 CSQFLRGQECVEECRVLQGLPREYVVARHCLCPCHPECOPOGNSVTCFGEADOCVACAHY 596
Db 531 CSQFLRGQECVEECRVLQGLPREYVVARHCLCPCHPECOPOGNSVTCFGEADOCVACAHY 590
Qy 597 KDPFFCVARCPGSKVDPDLSYMPWKFPDEGACQPCPCINCTHSCVDLDDKGCFAEORASP 656
Db 591 KDPFFCVARCPGSKVDPDLSYMPWKFPDEGACQPCPCINCTHSCVDLDDKGCFAEORASP 650
Qy 657 LTSIVSANGVILLVWLVGVVFGILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNQA 716
Db 651 LTSIIISAVGILLVWLVGVVFGILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNQA 710
Qy 717 QMRILKETELRKVKVLGSGAFVYKGIWIPDGENVKI PVAIKVLRNENTSPKANKEITLDE 776
Db 711 QMRILKETELRKVKVLGSGAFVYKGIWIPDGENVKI PVAIKVLRNENTSPKANKEITLDE 770
Qy 777 AYVMAGVGSYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRRGLSGDOLLNWCQIA 836
Db 771 AYVMAGVGSYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRRGLSGDOLLNWCQIA 830
Qy 837 KGMYSILEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMA 896
Db 831 KGMYSILEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMA 890
Qy 897 LESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGI PAREI PDLLEKGERLPOPICTID 956
Db 891 LESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGI PAREI PDLLEKGERLPOPICTID 950
Qy 957 VYIMVCKMIDSECRPRFRELVSFERSMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1016
Db 951 VYIMVCKMIDSECRPRFRELVSFERSMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDMGDLVDABEYLVPOQGFCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEA 1076
Db 1011 DDMGDLVDABEYLVPOQGFCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGDLGMGAAGLQSLPHTDPSPLQYSEDPTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGDLGMGAAGLQSLPHTDPSPLQYSEDPTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQEVYNQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVAFGGA 1196
Db 1131 LTCSPQEVYNQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVAFGGA 1190
Qy 1197 VENPEYLTPOGGAAPQHPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.

XX AC AAE24067;

XX DT 23-SEP-2002 (first entry)

XX DE Human Her-2 protein.

XX KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
XX KW hyperproliferative disorder; prophylaxis; inflammation; antisease;
XX KW tumour; gene therapy; phosphorothioate backbone.

XX OS Homo sapiens.

XX XX

PN WO200222636-A1.

XX

PD 21-MAR-2002 .
XX
XX 12-SEP-2001; 2001WO-US28572.
XX PR 15-SEP-2000; 2000US-0663834.
XX (ISIS-) ISIS PHARM INC.
XX PI Bennett CF, Cowsett LM;
XX WIPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX Example 13; Page 95-107; 116pp; English.
XX The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX SQ Sequence 1255 AA;

Query Match 98.0%; Score 6709; DB 23; Length 1255;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;
Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLLPASPTHLDMLRHLQYGCQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLLPASPTHLDMLRHLQYGCQVQGNL 60

Qy 61 ELTYLPTNASLSFLDIOEQVQYVLI AHNOVROVPLORLIRVGTQTFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLDIOEQVQYVLI AHNOVROVPLORLIRVGTQTFEDNYALAVLDNG 120

Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDF---NNFTV 177
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDF---NNFTV 177

Qy 178 SFWLRLVPKVSASHLE-NRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCGAGCARCKGP 236
Db 178 -----QLALTLDINRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCGAGCARCKGP 230

Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSHGICELHCPALVYNTTFFSMPNPEGRYT 296
Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSHGICELHCPALVYNTTFFSMPNPEGRYT 290

Qy 297 FGASCVTACPNYVLTVDVGSCTLVCPHLHNOEVTAEQDTCRCEKSKPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYVLTVDVGSCTLVCPHLHNOEVTAEQDTCRCEKSKPCARVCYGLGMEHL 350
Qy 357 REVRVAVTSANIQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLOVFTETLEITGYL 416
Db 351 REVRVAVTSANIQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLOVFTETLEITGYL 410
Qy 417 YISAWPDSLPDLVSFQNLQVIRGRILHNGAYSITLQGLIGISWGLRSLRSLGSLALIH 476
Db 411 YISAWPDSLPDLVSFQNLQVIRGRILHNGAYSITLQGLIGISWGLRSLRSLGSLALIH 470
Qy 477 NTHLCFVHTVPMDLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVN 536
Db 471 NTHLCFVHTVPMDLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVN 530
Qy 537 CSQFLRGQECVEECRVLQGLPREYVVARHCLCPCHPECOPOGNSVTCFGEADOCVACAHY 596

```
Db 531 CSQFLRQECVEECRVLOGLPREYVNAHCLUPCHPECOQPNQSVTCFGEADQCACAHY 590
Qy 597 KDPFFCVARCPGSKVPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS 656
Db 591 KDPFFCVARCPGSKVPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS 650
Qy 657 LTSIVSAVWGILLVVVGVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQA 716
Db 651 LTSIVSAVWGILLVVVGVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQA 710
Qy 717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDE 776
Db 711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDE 770
Qy 777 AYVMAGVSPYVSRLLGLCLTSTVQLVTQMLPYGCLLDHVRENRGRGLSQDOLLNWCQIA 836
Db 771 AYVMAGVSPYVSRLLGLCLTSTVQLVTQMLPYGCLLDHVRENRGRGLSQDOLLNWCQIA 830
Qy 837 KGSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVIKMA 896
Db 831 KGSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVIKMA 890
Qy 897 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 956
Db 891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 950
Qy 957 VYIMVWKCMIDSECRFRFRELVSFSEFMRMDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
Db 951 VYIMVWKCMIDSECRFRFRELVSFSEFMRMDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDMGDLVDAEYLVPQGFCDPDPAGAGGMVHRHRSSTRSGGDLTLGLPSEEEA 1076
Db 1011 DDMGDLVDAEYLVPQGFCDPDPAGAGGMVHRHRSSTRSGGDLTLGLPSEEEA 1070
Qy 1077 PRSLPASEGAGSDVFDGLGMAKGLQSLPTHDPSPLOYSDDTVPLPSETDGYVAP 1136
Db 1071 PRSLPASEGAGSDVFDGLGMAKGLQSLPTHDPSPLOYSDDTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYNQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFARFGA 1196
Db 1131 LTCSPQPEYNQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFARFGA 1190
Qy 1197 VENPEYLTPOGGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AC AAE20479;
XX DT 01-JUL-2002 (first entry)
XX DE Human Her-2/neu protein.
XX KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX OS human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX FT Homo sapiens.
XX FT Key Location/Qualifiers
XX FT Region 1021..1030
XX FT /note= "Naturally processed HLA-B44-restricted epitope"
XX PN WO200214503-A2.
```

```
PD 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
PF 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer -
XX Disclosure; Page 114-117; 129pp; English.
XX The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX Sequence 1255 AA;
```

```
Query Match 98.0%; Score 6709; DB 23; Length 1255;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLKRLPASPETHLDMRLHLYGQCQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLKRLPASPETHLDMRLHLYGQCQVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNQVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNQVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120

Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILWKDF---NNFTV 177
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILWKDF---NNFTV 177

Qy 178 SFWLVRPKVSAHLE-NRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGP 236
Db 178 -----QLALTLDITNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGP 230

Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRYT 290

Qy 297 FGASCVTACPNYLTSDVSGCTLVCPHNOEVAEDGTORCEKSKPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYLTSDVSGCTLVCPHNOEVAEDGTORCEKSKPCARVCYGLGMEHL 350

Qy 357 REVRVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQVFETIEETGYL 416
Db 357 REVRVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQVFETIEETGYL 416
```


QY 61 ELTYLPTNASLSFLQDIEVQGVVLIHNNVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIEVQGVVLIHNNVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDITLWKDF---NNFTV 177
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDITLWKDIFHKN--- 177
QY 178 SFMLRPVKVSASHLE-NRSRACHPCSPCKMGSRGWCSESSDCOSLTRTVCAGGCARCKGP 236
DB 178 -----QLALTLDITNRSRACHPCSPCKMGSRGWCSESSDCOSLTRTVCAGGCARCKGP 230
QY 237 LPTDCHEQCAACTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTTFESMPNPEGRYT 296
DB 231 LPTDCHEQCAACTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTTFESMPNPEGRYT 290
QY 297 FGASCVTACPNYLSLTDVSGCTLVCPHNOEVAEDGTQRCCKSPCARVCYGLGMEHL 356
DB 291 FGASCVTACPNYLSLTDVSGCTLVCPHNOEVAEDGTQRCCKSPCARVCYGLGMEHL 350
QY 357 REVRVTSANIQSFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYL 416
DB 351 REVRVTSANIQSFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYL 410
QY 417 YISAWPDSLPLDSVFQNLQVIRGRIILHNGAYSILTQGLGISWLGSLRLSGSLALIIHH 476
DB 411 YISAWPDSLPLDSVFQNLQVIRGRIILHNGAYSILTQGLGISWLGSLRLSGSLALIIHH 470
QY 477 NTHLCFVHTVPMDQLFRNPHQALLHTANRPEDSCVGEGLACHOLCARGHCWGPGTQCVN 536
DB 471 NTHLCFVHTVPMDQLFRNPHQALLHTANRPEDSCVGEGLACHOLCARGHCWGPGTQCVN 530
QY 537 CSQFLRGQECVEECRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADOCVACAHY 596
DB 531 CSQFLRGQECVEECRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADOCVACAHY 590
QY 597 KDPFPCVACPSGVKPDLSWMP1WKFPDEBACQPCPINCTHSCVDLDDKGCFAEGRASP 656
DB 591 KDPFPCVACPSGVKPDLSWMP1WKFPDEBACQPCPINCTHSCVDLDDKGCFAEGRASP 650
QY 657 LTSIVSAVVGILLVVLVGVVGLIKRRQOKIRKTYMRLLQETELVEPLTPSGAMPNQA 716
DB 651 LTSIVSAVVGILLVVLVGVVGLIKRRQOKIRKTYMRLLQETELVEPLTPSGAMPNQA 710
QY 717 QMRILKETELRKVKVLSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKELDE 776
DB 711 QMRILKETELRKVKVLSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKELDE 770
QY 777 AYVMAGVGSPPYVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENRLGSDLLNWCWQIA 836
DB 771 AYVMAGVGSPPYVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENRLGSDLLNWCWQIA 830
QY 837 KGMSYLEDVRLVHRDLAARNLVKSPNHVKITDPLGLARLLDIDETEVHADGGKVP1KWA 896
DB 831 KGMSYLEDVRLVHRDLAARNLVKSPNHVKITDPLGLARLLDIDETEVHADGGKVP1KWA 890
QY 897 LESILARRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 956
DB 891 LESILARRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950
QY 957 VYMIWVKWIMIDSECPRELVSEFSRMRDQRFVITQNEIDLGPASPLDSTFYRSLLE 1016
DB 951 VYMIWVKWIMIDSECPRELVSEFSRMRDQRFVITQNEIDLGPASPLDSTFYRSLLE 1010
QY 1017 DDDMGDLVDAEYLVPOQGFCCPDAPGAGGMVHRRHSSTSRSGGDLTLGLEPSEEA 1076
DB 1011 DDDMGDLVDAEYLVPOQGFCCPDAPGAGGMVHRRHSSTSRSGGDLTLGLEPSEEA 1070
QY 1077 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
DB 1071 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130
QY 1137 LTCSPQPEYVQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAGGA 1196

DB 1131 LTCSPOEYVQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAGGA 1190
QY 1197 VENPEYLTPOGGAPOPHPPAPSPAFDNLYYDQDPPERCAPPSTFKGTPTAENPEYLG 1256
DB 1191 VENPEYLTPOGGAPOPHPPAPSPAFDNLYYDQDPPERCAPPSTFKGTPTAENPEYLG 1250
QY 1257 LDVPV 1261
DB 1251 LDVPV 1255
RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX Human Her-2/neu polypeptide.
XX Human; Her-2/neu; cytostatic; haematological malignancy; CML;
XX acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
XX chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
XX Hodgkin's lymphoma; T cell therapy.
XX Homo sapiens.
XX
XX WO200213847-A2.
XX 21-FEB-2002.
XX 13-AUG-2001; 2001WO-US25408.
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
XX N-PSDB; ABK10730.
XX Inhibiting haematological malignancy development by administering
XX polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
XX encoding the polypeptide, or antigen presenting cells expressing the
XX polypeptide
XX
XX Disclosure; Page 71-74; 74pp; English.
XX
XX The invention relates to a method for inhibiting development of
XX haematological malignancy in a patient by administering a polypeptide
XX comprising an immunogenic portion of Her-2/neu or a polynucleotide
XX encoding the polypeptide. Antigen presenting cells that express the
XX protein can also be administered. The sequences are used for inhibiting
XX development of haematological malignancy such as acute myelogenous
XX leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
XX leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
XX lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
XX Sequence 1255 AA;
Query Match 98.0%; Score 6709; DB 23; Length 1255;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;
QY 1 MELALCRWGLLALLPPGAASVCTGDMKRLPASPETHLMLRHLHYGCGVQGNL 60
DB 1 MELALCRWGLLALLPPGAASVCTGDMKRLPASPETHLMLRHLHYGCGVQGNL 60
QY 61 ELTYLPTNASLSFLQDIEVQGVVLIHNNVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120

||||| 61 ELTYLPTNASLFLQDIQEVGYVLI AHNVQVRVPLQRLRIVRGTQOLFEDNYALAVLDNG 120
||||| 121 DPLNNTTPTVTCASPGGLRELQRLSITELTKGGVLIQRNPOLCYQDTILWKDF---NNFTV 177
||||| 121 DPLNNTTPTVTCASPGGLRELQRLSITELTKGGVLIQRNPOLCYQDTILWKDFIHKNN--- 177
||||| 178 SFWLVRPKVSASHLE-NRSRACHPCSPMKGSRGWGSESSDCQSLTRTVACAGGCARCKGP 236
||| : : : 178 -----QLALTLDITNRSRACHPCSPMKGSRGWGSESSDCQSLTRTVACAGGCARCKGP 230
||||| 237 LPTDCCHQCAAGCTGPKHSOCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYT 296
||||| 231 LPTDCCHQCAAGCTGPKHSOCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYT 290
||||| 297 FGASCVTACPNYLSLTDVSGCTLVCPHQNQVTAEDGTORCEKSKPCARVCYGLGWEHL 356
||||| 291 FGASCVTACPNYLSLTDVSGCTLVCPHQNQVTAEDGTORCEKSKPCARVCYGLGWEHL 350
||||| 357 REVRVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAFLOPEQLQVFTLEETGYL 416
||||| 351 REVRVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAFLOPEQLQVFTLEETGYL 410
||||| 417 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIH 476
||||| 411 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIH 470
||||| 477 NTHLCFVHTVPMDQLFRNPHQALLHTANRDECEVGEGLACHOLCARGHCGPPTQCVN 536
||||| 471 NTHLCFVHTVPMDQLFRNPHQALLHTANRDECEVGEGLACHOLCARGHCGPPTQCVN 530
||||| 537 CSQFLRGQCEVEECRVLQGLPREYVYARHCLPCHPECPQNGSVTCFGEADQCVACAHY 596
||||| 531 CSQFLRGQCEVEECRVLQGLPREYVYARHCLPCHPECPQNGSVTCFGEADQCVACAHY 590
||||| 597 KDPFPCVARGSVKVPDLSYMPIWKFDEEGACQPCINCHSCVDLDDKCPABEQRAS 656
||||| 591 KDPFPCVARGSVKVPDLSYMPIWKFDEEGACQPCINCHSCVDLDDKCPABEQRAS 650
||||| 657 LTSIVSAVVGILLVVLGWFGLIKRQKIRKYTMRLLOQETELVEPLTPSGAMPNOA 716
||||| 651 LTSIVSAVVGILLVVLGWFGLIKRQKIRKYTMRLLOQETELVEPLTPSGAMPNOA 710
||||| 717 QMRILKETELRKVKVLSGAFGTGVYGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDE 776
||||| 711 QMRILKETELRKVKVLSGAFGTGVYGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDE 770
||||| 777 AYVMAGVSPYVSRLLGLCLTSTVOLVTLQMPYGCCLLDHVRENRGRIGSQDLNWCQIA 836
||||| 771 AYVMAGVSPYVSRLLGLCLTSTVOLVTLQMPYGCCLLDHVRENRGRIGSQDLNWCQIA 830
||||| 837 KGMVLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKAWA 896
||||| 831 KGMVLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKAWA 890
||||| 897 LESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 956
||||| 891 LESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 950
||||| 957 VYMTMVKWMTDSECRPRFRELSEFSESRMARDPQRFVVIQNEGLGPASPLDSTFVRSLE 1016
||||| 951 VYMTMVKWMTDSECRPRFRELSEFSESRMARDPQRFVVIQNEGLGPASPLDSTFVRSLE 1010
||||| 1017 DDDMGDLVDAAEYLVLPQGGFCPPDAPAGAGMWHHRSSSTRSGGGDLTLGLPSEBEA 1076
||||| 1011 DDDMGDLVDAAEYLVLPQGGFCPPDAPAGAGMWHHRSSSTRSGGGDLTLGLPSEBEA 1070
||||| 1077 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAP 1136
||||| 1071 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAP 1130
||||| 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAGGA 1196
|||||

Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAGGA 1190
Qy 1197 VENPEYLTPOGGAAPQPHPPAFSPAFDNLYWQDPPERGAPPSTFKGTPTAENPEYLG 1256
||||| 1191 VENPEYLTPOGGAAPQPHPPAFSPAFDNLYWQDPPERGAPPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
||||| 1251 LDVPV 1255
Db
RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX AAR39568;
AC AAR39568;
XX 07-FEB-1994 (first entry)
XX Sequence of c-erbB-2 tumour antigen.
DE Tumour antigen; c-erbB-2; glycoprotein.
XX Homo sapiens.
OS WO9316185-A.
PN 19-AUG-1993.
PD 05-FEB-1993; 93WO-US01055.
PF 06-FEB-1992; 92US-0831967.
PR (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX Houston Lb, Huston JS, Oppermann H, Ring DB;
PI WPI; 1993-272889/34.
DR N-PSDB; AAQ46083.
XX New single chain Fv polypeptide binding to C-erbB-2 tumour antigen - for imaging or treating breast or ovarian cancer etc.
PT Disclosure; pages 48-54; 87pp; English.
PS c-erbB-2 refers to a protein antigen expressed on the surface of tumour cells, such as breast and ovarian tumour cells, which is an approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;
Query Match 97.3%; Score 6666; DB 14; Length 1433;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1233; Conservative 9; Mismatches 9; Indels 14; Gaps 3;
Qy 1 MELAALCRWGLLLALLPFGAASTQVCTGDMKRLPASPETHDMLRHLGYQCVVQGNL 60
||||| 1 MELAALCRWGLLLALLPFGAASTQVCTGDMKRLPASPETHDMLRHLGYQCVVQGNL 60
Db 61 ELTYLPTNASLFLQDIQEVGYVLI AHNVQVRVPLQRLRIVRGTQOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIQEVGYVLI AHNVQVRVPLQRLRIVRGTQOLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSITELTKGGVLIQRNPOLCYQDTILWKDF---NNFTV 177
||||| 121 DPLNNTTPTVTCASPGGLRELQRLSITELTKGGVLIQRNPOLCYQDTILWKDFIHKNN--- 177
Db 178 SFWLVRPKVSASHLE-NRSRACHPCSPMKGSRGWGSESSDCQSLTRTVACAGGCARCKGP 236
||| : : : 178 -----QLALTLDITNRSRACHPCSPMKGSRGWGSESSDCQSLTRTVACAGGCARCKGP 230
Db

Search completed: July 22, 2003, 08:40:44
Job time : 43.1589 secs

| | | | |
|----|------|---|------|
| Qy | 237 | LPTDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYT | 296 |
| Db | 231 | LPTDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYT | 290 |
| Qy | 297 | FGASCVTACPYNYLSTDVGSCTLVCPHNOEVTAEQCEKCKPCARVCYGLGWEHL | 356 |
| Db | 291 | FGASCVTACPYNYLSTDVGSCTLVCPHNOEVTAEQCEKCKPCARVCYGLGWEHL | 350 |
| Qy | 357 | REVRVTSANIQEFACKXIFGSLAFPEFSGDGPASNTAPLQPEQLQVFETLEEITGYL | 416 |
| Db | 351 | REVRVTSANIQEFACKXIFGSLAFPEFSGDGPASNTAPLQPEHLQVFETLEEITGYL | 410 |
| Qy | 417 | YISAWPDSLPLSVFQNLQVIRGRILHNGAYSITLQGLGSIWGLRSLRELGSGLAIHH | 476 |
| Db | 411 | YISAWPDSLPLSVFQNLQVIRGRILHNGAYSITLQGLGSIWGLRSLRELGSGLAIHH | 470 |
| Qy | 477 | NTHLCFVHTVPWDOLFNRPHQALLHTANRPEDECVGEGLACHOLCARGHGWGPQTQCVN | 536 |
| Db | 471 | NTHLSFVHTVPWDOLFNRPHQALLHTANRPEDECVGEGLACHOLCARGHGWGPQTQCVN | 530 |
| Qy | 537 | CSQFLRGQCBECRVLQGLPREYVNAHCLPCHPECQFQNGSVTCFGEADQCVACAHY | 596 |
| Db | 531 | CSQFLRGQCBECRVLQGLPREYVNAHCLPCHPECQFQNGSVTCFGEADQCVACAHY | 590 |
| Qy | 597 | KOPPFCAVCPGSKVXPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEABORASP | 656 |
| Db | 591 | KOPPFCAVCPGSKVXPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEABORASP | 650 |
| Qy | 657 | LTSISAVVGILLVVVLGVVFGILIKRROQKIRKYMRRLLQETELVEPLTPSGAMPNOA | 716 |
| Db | 651 | LTSISAVVGILLVVVLGVVFGILIKRROQKIRKYMRRLLQETELVEPLTPSGAMPNOA | 710 |
| Qy | 717 | QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPLANKEILDE | 776 |
| Db | 711 | QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPLANKEILDE | 770 |
| Qy | 777 | AYVMAGVGSPPYVSRLLIGICLTSTVQLVTQMLPYGCLLDHVRENRRGLGSODLLNWCQIA | 836 |
| Db | 771 | AYVMAGVGSPPYVSRLLIGICLTSTVQLVTQMLPYGCLLDHVRENRRGLGSODLLNWCQIA | 830 |
| Qy | 837 | KGMSYLEDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPKMA | 896 |
| Db | 831 | KGMSYLEDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPKMA | 890 |
| Qy | 897 | LESILRRRTHQSDVMSYGVTVWELMTFCAKPYDGIPAREIPDLLEKGERLPQPPICTID | 956 |
| Db | 891 | LESILRRRTHQSDVMSYGVTVWELMTFCAKPYDGIPAREIPDLLEKGERLPQPPICTID | 950 |
| Qy | 957 | VYMIWVKWMDSECRPRELVSERSMARDPQRFVITQNEGLGPASPLDSTFYRSLLE | 1016 |
| Db | 951 | VYMIWVKWMDSECRPRELVSERSMARDPQRFVITQNEGLGPASPLDSTFYRSLLE | 1010 |
| Qy | 1017 | DDDMGDLVDAEYLVPOQGFCCPDAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEA | 1076 |
| Db | 1011 | DDDMGDLVDAEYLVPOQGFCCPDAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEA | 1070 |
| Qy | 1077 | PRSLAPSEGAGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAP | 1136 |
| Db | 1071 | PRSLAPSEGAGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAP | 1130 |
| Qy | 1137 | LTCSPQEVYNQPDVVRPQPPSPREGPLPAARAGATLERAKTLPSPKNGVVKDVFAGGA | 1196 |
| Db | 1131 | LTCSPQEVYNQPDVVRPQPPSPREGPLPAARAGATLERAKTLPSPKNGVVKDVFAGGA | 1190 |
| Qy | 1197 | VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG | 1256 |
| Db | 1191 | VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG | 1250 |
| Qy | 1257 | LDVPV 1261 | |
| Db | 1251 | LDVPV 1255 | |

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-210-224-14
Perfect score: 6827
Sequence: 1 MELALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:*

1: Pir1: *
2: Pir2: *
3: Pir3: *
4: Pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 6684 | 97.9 | 1255 | 1 A24571 | protein-tyrosine k |
| 2 | 5870.5 | 86.0 | 1254 | 2 I48161 | p-185 precursor - |
| 3 | 5866 | 85.9 | 1260 | 1 TVRTNU | protein-tyrosine k |
| 4 | 3054 | 44.7 | 1210 | 1 GOHUE | epidermal growth f |
| 5 | 3030 | 44.4 | 1210 | 2 A53183 | epidermal growth f |
| 6 | 3009.5 | 44.1 | 1223 | 1 TVCHLV | epidermal growth f |
| 7 | 2905.5 | 42.6 | 1308 | 2 A47253 | epidermal growth f |
| 8 | 2602 | 38.1 | 1166 | 1 S06142 | protein-tyrosine k |
| 9 | 2333.5 | 34.2 | 1342 | 2 A36223 | kinase-related tra |
| 10 | 2248.5 | 32.9 | 1339 | 2 JC4387 | epidermal growth f |
| 11 | 1766.5 | 25.9 | 698 | 1 TVFVLV | protein-tyrosine k |
| 12 | 1703 | 24.9 | 604 | 1 TVYUHV | protein-tyrosine k |
| 13 | 1647 | 24.1 | 544 | 2 S35745 | protein-tyrosine k |
| 14 | 1640 | 23.0 | 545 | 2 S00727 | protein-tyrosine k |
| 15 | 1623 | 23.8 | 540 | 2 B44776 | kinase-related tra |
| 16 | 1621 | 23.7 | 540 | 1 TVFVEB | protein-tyrosine k |
| 17 | 1561.5 | 22.9 | 1330 | 1 GQFFE | protein-tyrosine k |
| 18 | 1432 | 20.8 | 644 | 2 A36325 | epidermal growth f |
| 19 | 1251 | 18.3 | 1323 | 2 E88257 | epidermal growth f |
| 20 | 1251 | 18.3 | 1374 | 2 S07012 | protein let-23 [im |
| 21 | 1171 | 17.2 | 1369 | 2 S07013 | protein-tyrosine k |
| 22 | 1142 | 16.7 | 1717 | 1 A45558 | epidermal growth f |
| 23 | 1041 | 15.2 | 527 | 2 A42032 | epidermal growth f |
| 24 | 906.5 | 13.3 | 843 | 2 A21731 | epidermal growth f |
| 25 | 806.5 | 11.8 | 346 | 2 S13807 | protein-tyrosine k |
| 26 | 754.5 | 11.1 | 311 | 2 S13808 | protein-tyrosine k |
| 27 | 680 | 10.0 | 1382 | 1 INHUR | insulin receptor p |
| 28 | 671 | 9.8 | 1363 | 2 T43220 | insulin-like growt |
| 29 | 671 | 9.8 | 1383 | 2 A36080 | insulin receptor p |

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein e
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, I.
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epic
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SEM>
A;Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282
R;Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chr
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:G183986
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcripti
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

ALIGNMENTS

30 670.5 9.8 1372 2 A34157 insulin receptor p
31 661 9.7 1607 2 T43212 insulin-like growt
32 655.5 9.6 1300 2 A36502 insulin receptor-r
33 644 9.4 1268 2 B36502 insulin receptor-r
34 643 9.4 1477 2 T18534 protein-tyrosine k
35 604 8.8 1367 1 IGHUR1 insulin-like growt
36 596 8.7 1371 2 A33837 insulin-like growt
37 591 8.7 987 2 A54092 protein-tyrosine k
38 589.5 8.6 977 2 S49004 tyrosine kinase Mp
39 588 8.6 1114 1 S05582 protein-tyrosine k
40 588 8.6 1390 2 T30346 insulin receptor -
41 586 8.6 1091 2 S33596 protein-tyrosine k
42 585.5 8.6 2148 1 A56081 insulin receptor -
43 584.5 8.6 976 2 A36355 protein-tyrosine k
44 583.5 8.5 952 2 I50612 protein-tyrosine k
45 583 8.5 987 2 I48652 mouse developmenta

A;Cross-references: GB:M16792; NID:g183903; PIDN:AAAS8637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
kinase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-125/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F:22-653/Domain: extracellular #status predicted <EXT>
F:70-304/Domain: EGF receptor extracellular domain repeat <BE1>
F:395-605/Domain: EGF receptor extracellular domain repeat <BE2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
F:68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.9%; Score 6684; DB 1; Length 1255;
Best Local Similarity 97.6%; Pred. No. 3.1e-279;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASCTGCTDKMLRLPASPETHLDMRLHLYGCGVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASCTGCTDKMLRLPASPETHLDMRLHLYGCGVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVGYVLIHANNVQVPLQRLRVGTQTLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEVGYVLIHANNVQVPLQRLRVGTQTLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGLREQLRLSLTEILKGGVLIQNPQLCYQDTILWKDIFHNQOLA 180
DB 121 DPLNNTTPTVGTASPGGLREQLRLSLTEILKGGVLIQNPQLCYQDTILWKDIFHNQOLA 180
QY 181 LTLIDTNRSRACHPCSPMKGSRGCGESSEDCQSLTRTVACGAGCARCKGPLPNNFTVSF 240
DB 181 LTLIDTNRSRACHPCSPMKGSRGCGESSEDCQSLTRTVACGAGCARCKGPLP----- 232
QY 241 WLRVPKVSASHLE-----KHSOCLACLHFNHSGICELHCPALVTYNTDTFESMPNPE 292
DB 233 -----TDCCEQAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPE 286
QY 293 GRYTFGASCVTACPVNYLSTDVGSLVCPPLHNOEVTAEADGTCRCEKSPCARVCYGLG 352
DB 287 GRYTFGASCVTACPVNYLSTDVGSLVCPPLHNOEVTAEADGTCRCEKSPCARVCYGLG 346
QY 353 MEHLREVRVTSANIQEFAGCKTIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEEI 412
DB 347 MEHLREVRVTSANIQEFAGCKTIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEEI 406
QY 413 TGYLYISAWPDSLPLDSVFNQLQVIRGRILHNGAYSILTLQGLIGISWGLRLSRLGSLA 472
DB 407 TGYLYISAWPDSLPLDSVFNQLQVIRGRILHNGAYSILTLQGLIGISWGLRLSRLGSLA 466
QY 473 LIHNTHLCFVHTVPMQDLPNPHQALLHTANRDEECVCEGLACHOLCARGHCWGPGPT 532
DB 467 LIHNTHLCFVHTVPMQDLPNPHQALLHTANRDEECVCEGLACHOLCARGHCWGPGPT 526
QY 533 QCVNCSQFLRGQECVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQVCA 592
DB 527 QCVNCSQFLRGQECVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQVCA 586
QY 593 CAHYKDPFPFCVARCPGKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 652

DB 587 CAHYKDPFPFCVARCPGKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
QY 653 RASPLTSIVSNAVVCILLVVLGVVFGVGLIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 712
DB 647 RASPLTSIIISAVGILLVVLGVVFGVGLIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 706
QY 713 PNOAQMRLKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 772
DB 707 PNOAQMRLKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 766
QY 773 ILDBAYMAGVSGPYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRRGLGSDLLNWC 832
DB 767 ILDBAYMAGVSGPYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRRGLGSDLLNWC 826
QY 833 MOTAKGMSYLEDVRLVHRDLAARNVLKSPNHNKTIQFGLARLLDIDETEHADGGKVP 892
DB 827 MOTAKGMSYLEDVRLVHRDLAARNVLKSPNHNKTIQFGLARLLDIDETEHADGGKVP 886
QY 893 KMALESILRRRFTHQSDVMSYGVYVWELMTFGAKPYDGIPIAREIPDILLEKGERLPQPP 952
DB 887 KMALESILRRRFTHQSDVMSYGVYVWELMTFGAKPYDGIPIAREIPDILLEKGERLPQPP 946
QY 953 CTIDVYMIMVKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYR 1012
DB 947 CTIDVYMIMVKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYR 1006
QY 1013 SLLEDDDMGDLVDAEEYLVPOQGFCCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPS 1072
DB 1007 SLLEDDDMGDLVDAEEYLVPOQGFCCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPS 1066
QY 1073 EEEAPRSLAPSEGAGSDVFDGDLGMAAGKQLSLPTHDPSPLOQYSEDPTVPLPSETDG 1132
DB 1067 EEEAPRSLAPSEGAGSDVFDGDLGMAAGKQLSLPTHDPSPLOQYSEDPTVPLPSETDG 1126
QY 1133 YVAPLTCSPQPEYVYNQPDVPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVA 1192
DB 1127 YVAPLTCSPQPEYVYNQPDVPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVA 1186
QY 1193 FGGAVENPEYLTPOGGAAPOPHPPAPAFSPADNLYYNDODPERGAPSTFKGTPTAENP 1252
DB 1187 FGGAVENPEYLTPOGGAAPOPHPPAPAFSPADNLYYNDODPERGAPSTFKGTPTAENP 1246
QY 1253 EYLGLDVVP 1261
DB 1247 EYLGLDVVP 1255

RESULT 2

148161

p-185 precursor - golden hamster

C;Species: Mesocricetus auratus (golden hamster)

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999

C;Accession: 148161

R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Yamazaki, Y.; Ishikawa, T.

Gene 140, 251-255, 1994

A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.

A;Reference number: 148161; MUID:94193007; PMID:7908275

A;Accession: 148161

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1254 <RES>

A;Cross-references: GB:D16295; NID:9493236; PIDN:BAA03801.1; PID:g747595

C;Genetics:

A;Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology

F;Keywords: ATP

F;718-983/Domain: protein kinase homology <KIN>

F;726-734/Region: protein kinase ATP-binding motif

Query Match 86.0%; Score 5870.5; DB 2; Length 1254;

Best Local Similarity 84.9%; Pred. No. 2.2e-244;

Matches 1088; Conservative 57; Mismatches 89; Indels 47; Gaps 4;

```
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60
Dy 1 MELAAWCGWGLLLALLSPGASCTGCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVGYVLIAHNQVQVPLQRLRVIRGTQFEDNYALAVLNG 120
Dy 61 ELTYLPTANATLSFLQDIOEVGYVLIAHNSVHRHVPQLRLRVIRGTQFEDKYALAVLNR 120
Qy 121 DPLNNTTPTVTCASPGGLREQLRLSLTEILKGGVLIQRNPQLCYODTILWKDIFHNKQOLA 180
Dy 121 DPLDNVTATGRTPEGLREQLRLSLTEILKGGVLIQRNPQLCYODTILWKDIFRKNQOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGSRKWSSESDECSLRTV-----C 220
Dy 181 PVDIDTNRSRACPPCAPACKDNHCWASPEDCQTLTGTIAPRAVPAARLARLTDCCHBOC 240
Qy 221 AGGCARCKGPLPFPNNFTVSFWLRVPKVSASHLEKHSCLACLHFNHSGICELHCPALVTY 280
Dy 241 AAGCT---GP-----KHSCLACLHFNHSGICELHCPALVTY 274
Qy 281 NTDTFESMPNPEGRYTFGASCVTACPNYLVSTDVSGSCTLVCPLNHQNVEATSDGTQRCCK 340
Dy 275 NTDTFESMPNPEGRYTFGASCVTTCPNYLVSTEVSGSCTLVCPLNHQNVEATSDGTQRCCK 334
Qy 341 SKPCARVCYGLGMEHLREVRAVTSANTQEFAGCKKIFGSLAFLESFDDGPASNTAPLQP 400
Dy 335 SKSCARVCYGLGMEHLRGARAITSANIQEFAGCKKIFGSLAFLESFDDGNSSGIAPLTP 394
Qy 401 EQLQVFTELEITGVLYISAWPDSI.PDLSVFQNLQVIRGRILHNGAYSLTQGLGISHLG 460
Dy 395 EQLQVFTELEITGVLYISAWPDSLDLSVFQNLQVIRGRVLRHLDGAYSLAQGLGIRWLG 454
Qy 461 LRSRELGSGLALIHNTLHLCFVHTVPMQDLFRNPHQALLHTANRPEDECVBEGGLACHOL 520
Dy 455 LRSRELGSGLVLIHRNTHLCFVHTVPMQDLFRNPHQALLHSGNSEECGLKDFACVPL 514
Qy 521 CARGHCWGPGTQCNSQFIRGQECVEECVRLQGLPREYNARHCLPCHPECPQNGSV 580
Dy 515 CAHGHCGPGTQCNSHFIRGQECVKECRWKGLPREYVNGKHLPCHPCECPQNSTE 574
Qy 581 TCFPEADQCVACAHYKDPFCVARGCPGVKPDLSYMPIWKFPDEEGACQPCINCHTSC 640
Dy 575 TCTGSEADQCTACPHYKDSPPCVARGCPGVKPDLSYMPIWKYPDEEGMCQPCINCHTSC 634
Qy 641 VLDLDDKCPAQRASPLTSIVSAVVGLLVVGLVGVFGLIKRRQKIRKTYMRLLQET 700
Dy 635 VLDLDDKCPAQRASPLTSIIATVVGILLFLVIGVVGILLIKRRQKIRKTYMRLLQET 694
Qy 701 ELVEPLTPSGAMPNQAQNRILKETELRKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKV 760
Dy 695 ELVEPLTPSGAMPNQAQNRILKETELRKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKV 754
Qy 761 LRENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVTLQMPYGCGLLDHVRNR 820
Dy 755 LRENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVTLQMPYGCGLLDHVRNR 814
Qy 821 GRLGSDLLNCMQIAKGMSYLEVRLVHRDLAARNVLKSPNPKVITDFGLARLLDIDE 880
Dy 815 GRLGSDLLNCWQVIAKGMSYLEVRLVHRDLAARNVLKSPNPKVITDFGLARLLDIDE 874
Qy 881 TEHADCGKPIKMALESILRRRFTQSDVWSYGVTVWELMTGAKPYDGIPIAREIPDL 940
Dy 875 TEHADCGKPIKMALESILRRRFTQSDVWSYGVTVWELMTGAKPYDGIPIAREIPDL 934
Qy 941 LEKERLUPQPPICITDVMIMVKCWMIDSECRPFRELVSEFSRWARDPQRFVIONEDL 1000
Dy 935 LEKERLUPQPPICITDVMIMVKCWMIDSECRPFRELVSEFSRWARDPQRFVIONEDL 994
Qy 1001 GPASPLDSTFYRSLLEDMDGLVDABEYLVPOQGFCCPDPAAGAGGVVHRRHSSSTRS 1060
Dy 995 GPSSPLDSTFYRSLLEDMDGLVDABEYLVPOQGFCCPDPAAGAGSTAHRHSSSTRS 1054
Qy 1061 GGGDLTLGLEPSEEEAPRSLAPSEGASDVFDGDLGMAKGLQSLTPTHDPSPLQRYSE 1120
```

```
Db 1055 GGGELTLGMEPSGEEPPRSLAPSEGAGSVFEGELMGATKGPQISPRDLSPLORYSE 1114
Qy 1121 DPTVPLSETGYVAPLTCSPQPEYVQDPVVRPQSPREGPLPAARPAAGATLERAKTUS 1180
Dy 1115 DPTLPLPTETDGYVAPLACSPQPEYVQDPVVRPQSPPLTEGGLPPVRPAGATLERPKTUS 1174
Qy 1181 PGKGVKVDVFAFGGAVENPEYLTPQGAAAPQHPHPPAFSPAFDNLXYWQDPPPPGAPP 1240
Dy 1175 PGKGVKVDVTFGGAVENPEYLVPRGGSASQPH-PPALCPAFDNLXYWQDPPSERGSP 1233
Qy 1241 STFKGTPTAENPEYGLGLDVPV 1261
Dy 1234 NTFEGTPTAENPEYGLGLDVPV 1254

RESULT 3
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1998 #sequence_revision 31-Dec-1998 #text_change 11-Jun-1999
C:Accession: A24562; A61204
R:Barthmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein
A:Reference number: A24562; MUID:86118662; PMID:3945311
A:Accession: A24562
A:Molecule type: mRNA
A:Residues: 1-1260 <BAR>
A:Cross-references: EMBL:X03362; NID:G56745; PIDN:CAA27059.1; PID:G56746
R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,
Carlinogenesis 12, 1975-1978, 1991
A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals a
2-thiazolylformamide or N-methyl-N-nitrosourea.
A:Reference number: A61204; MUID:92035293; PMID:1682063
A:Accession: A61204
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 637-663,'V',665-702 <MAS>
A:Note: authors translated the codon GCA for residue 25 as Val
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; pho;
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase new #status predicted <MAT>
F:658-680/Domain: transmembrane #status predicted <TMN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carboxylate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.9%; Score 5866; DB 1; Length 1260;
Best Local Similarity 85.7%; Pred. No. 3,5e-244;
Matches 1089; Conservative 51; Mismatches 107; Indels 24; Gaps 4;
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60
Dy 4 MELAAWCGWGLLLALLSPGASCTGCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 63
Qy 61 ELTYLPTNASLSFLQDIOEVGYVLIAHNQVQVPLQRLRVIRGTQFEDNYALAVLNG 120
Dy 64 ELTYVPAANASLSFLQDIOEVGYVLIAHNQVQVPLQRLRVIRGTQFEDKYALAVLNR 123
Qy 121 DPLNNTTPTVTCASPGGLREQLRLSLTEILKGGVLIQRNPQLCYODTILWKDIFHNKQOLA 179
Dy 124 DPQDNVAASPTGRTPEGLREQLRLSLTEILKGGVLIQRNPQLCYODVWLKDFRKNQOLA 183
Qy 180 ALTLIDTNRSRACHPCSPMKSGSRKWSSESDECSLRTVTCAGGACRCKGPLPFPNNFTVS 239
Dy 184 APVIDTNRSRACPPCAPACKDNHCWGESPEDCQILGTICTSCARCKGRPL----- 236
```


R; Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A; Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor.
A; Reference number: A60143; MUID: 85182650; PMID: 2985580
A; Molecule type: protein
A; Residues: 740-744, X', 746-747 <RUS>
R; Mroczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A; Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase.
A; Reference number: A38023; MUID: 94191554; PMID: 6325948
A; Comments: annotation: receptor activity
A; Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R; Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
Cell 59, 33-43, 1989
A; Title: Functional independence of the epidermal growth factor receptor from a domain in the cytoplasmic tail.
A; Reference number: A33311; MUID: 90003233; PMID: 2790960
A; Comments: annotation: internalization signal
C; Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex.
C; Genes: EGFR
A; Gene: GDB: 120610; OMIM: 131550
A; Cross-references: GDB: 120610; OMIM: 131550
A; Map position: 7p12.3-7p12.1
C; Superfamily: epidermal growth factor receptor; protein kinase homology
C; Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphotyrosine
F; 1-24/Domain: signal sequence #status predicted <SIG>
F; 25-1210/Product: EGF receptor #status predicted <MAT>
F; 25-645/Domain: extracellular #status predicted <EXT>
F; 75-300/Domain: EGF receptor extracellular domain repeat <EB1>
F; 390-600/Domain: EGF receptor extracellular domain repeat <EB2>
F; 646-668/Domain: transmembrane #status predicted <TM>
F; 669-1210/Domain: intracellular #status predicted <INT>
F; 710-975/Domain: protein kinase homology <KIN>
F; 718-726/Region: protein kinase ATP-binding motif
F; 999-1046/Region: coated-pit mediated internalization signal
F; 1047-1210/Region: inhibitory
F; 128, 175, 352, 413, 444, 528, 603/Binding site: carboxylate (Asn) (covalent) #status predicted
F; 745/Active site: Lys #status experimental

Query Match 44.7%; Score 3054; DB 1; Length 1210;
Best Local Similarity 48.3%; Pred. No. 9.8e-124;
Matches 618; Conservative 179; Mismatches 354; Indels 128; Gaps 23;

QY 11 LLLALLPPGNA--STQVCTGTDMLRLPASPETHLDMLRLHYGCGVQVQGNLELTYPTN 68
DB 14 LLAALCPASRALEEKVKCOGTSNKLTLQGLTFEDFLSLQRMFNCEVNLGEBITYQVN 73
QY 69 ASLSFLQDIQEVQYVLIHNOVRQVPLQRLRVGTQLFEDNYALAVLDNGDPLNNTTP 128
DB 74 YDLSFLTKIQEVAGYVLIATNVERIPLENLQIRGNMYENSYALAVLSNYD----- 126
QY 129 VTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNOLATLIDTNR 188
DB 127 --ANKTGLKELPMRLQELTHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFNH 193
QY 189 SRACHPCSPKCSRCWGESSEDCSLTRTVAGGCA-RCKGPLPNNFTVSWLRVPKV 247
DB 184 LSCQCKDCSPNGSCWGAAGEENCQKLTIIICAOQCSGRCKRGSFSD----- 230
QY 248 SASHLF-----KHSDCLACLHFNHSGICELRCPALVTYNTDTFESMPNPEGRYTFGA 299
DB 231 -CHNQCAAGCTCPRESDCLVCKRFRDEATCKDTCPLMLYNFTTYQMDVNPGRKYSFGA 289
QY 300 SCVTACPYNYLSTVSGCTLVCPHNLHNOVTAEDGTORCEKCKSPCARVCYGLGMEHLREV 359
DB 290 TCVKKCPNRYVVDHGSVCVRACGADSYEM-EEDGVKCKKCEGCPCKVNGVIGIGEFKDS 348
QY 360 RAVTSANIQEFAGCKIFGSLAFPLPESFGDPSASNTAPLOEQLVFTELEETGYLYIS 419
DB 349 LSNATNKHFKNCTSIISGDLHILPVAFRGDSFTHTPPLDPQELDLTKVETLTGELLIQ 408
QY 420 ANPDSLPLDLSVFQNLQVIRGRILHNGAYSILTLQGLIGSWLGRSLRELGSGLALIHNTH 479

409 ANPENTDHLAFENLEIRGRITQHQGFSLAVVSLNITSLGLRSLKSIDSGDVIIISGNKN 468
QY 480 LCFVHTVPMQDLFRNPHQALLHTANRDEDECVGSEGLACHOLCARGHGWPGPOTOCVNCQ 539
DB 469 LCYANTINWKKLFGTSQCKTKIISNRGENSKATGQVCHALCSPGCGNGPEPCVSCRN 528
QY 540 FLRGQECVBEICVQLGLPREYVNAHCLPCHPCEQPONGSVTCFGEADQCACAHYKDP 599
DB 529 VSRGECVCKKLEGEPRFVENSECICQHPCELPQAMNITCTGRGPDNCIOCAHYIDG 588
QY 600 PFCVACPCSVKPDLSVMPYKWPDEGACQPCINCTHSCVDLDDKCPAEQASPLTS 659
DB 589 PHCVKTCPCAGVMGENNLT-VKMYADAGHVCHLCPNCTYCTGPGLEGCTNGPKIP--S 645
QY 660 IVSAVVG---ILVWVGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQA 716
DB 646 IATGMVGAALLLVVALGIG---LFMRHHIVKRLTLRLLOERELVEPLTPSGEAPNQA 702
QY 717 QMRILKETELRKVVLGSGAFVTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDE 776
DB 703 LLRIKETEFKIKVLGSGAFVTVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDE 762
QY 777 AYVMAGVSPYVSRLLGICLTSTVOLVTLQMPYGCCLLDHYRENRRGLSGDOLLNMCQIA 836
DB 763 AYVMASVDNPHVCRLLGICLTSTVOLITQMPFGCLLDYVREHKDNTGSOYLLNMCQIA 822
QY 837 KGMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMA 896
DB 823 KGMNILEDRLVHRDLAARNVVKTPQHVKITDFGLAKLLGAEBEYHAEKGKVPKIKMA 882
QY 897 LESILRRRPTHQSDVMSYGVTVNMLMTFGAKPYDGPAREIPDLLEKGERLPQPPICITD 956
DB 883 LESILHRIYTHQSDVMSYGVTVNMLMTFGSKPYDGPAREIPDLLEKGERLPQPPICITD 942
QY 957 VYMIWVKCMIDSECEPRRELVSERMAARDQRFVVIQ-NEDLGPASPLDSTFYRSL 1015
DB 943 VYMIWVKCMIDSECEPRRELVSERMAARDQRFVVIQ-NEDLGPASPLDSTFYRSL 1002
QY 1016 EDDMDGDLVDAEYLVPOQGFPCDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEE 1075
DB 1003 DEEDMDVDVDAEYLVPOQGF-----SPPS 1028
QY 1076 APRSLAPSEGAGSDVFDGLGMAAGKLSLTHDPSPQRYSEDPTVPLPSET--DGY 1133
DB 1029 TSRTPLLSLSATS--NSTVACIDRNLQSCPIKEDSFQRYSSDPTGALTEDSIDDTF 1086
QY 1134 VAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVPF 1193
DB 1087 L-----PVPEYINQ-SVPKRPAGSVQNPVYHQNPLN-----APSRDPHYQD--PH 1129
QY 1194 GGAVENPEYL-TPOGGAAPQHPHPPAFSPAFDNLVYWDQ-----DP-----PER 1236
DB 1130 STAVGNPEYLVNTVQ-----PTCVNSTFDSFAHWAQKSHQISLDNPDYQDFFPKE 1180
QY 1237 GAPPSTFKGTPTAENPEYL 1255
DB 1181 AKPNGIFKGS-TAENAEYL 1198

RESULT 5
A53183
epidermal growth factor receptor precursor - mouse
C; Species: Mus musculus (house mouse)
C; Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
C; Accession: A53183; A31818; S24942; A28941; S45325; I49643
R; Luetcke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.F.
Genes Dev. 8, 399-413, 1994
A; Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor.
A; Reference number: A53183; MUID: 94170986; PMID: 8125255
A; Accession: A53183
A; Molecule type: mRNA
A; Residues: 1-1210 <LUE>
A; Cross-references: GB: U03425


```

RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Gallus gallus (chicken)
C>Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720; A00643
R:Lab, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennart
Mol. Cell. Biol. 8, 1970-1978, 1988
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mod
A:Reference number: A27720; MUID:88261272; PMID:3260329
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carboxylate (Thr) (covalent) #E
F:192,650/Binding site: carboxylate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 44.1%; Score 3009.5; DB 1; Length 1223;
Best Local Similarity 47.3%; Pred. No. 86-122;
Matches 620; Conservative 176; Mismatches 348; Indels 167; Gaps 27;
QY 8 RWGLLALLPPCAA-----STQVCTGTDMLRLPASPEHLDMLRLHYQGCVQGNLE 61
DB 13 RGAALVLLLLGVALCSAVEEKKVCGTNNKLTQLGHVEDHFTSLQRMVNCVLSNLE 72
QY 62 LTYLPTNASLFDIQEYQGVLYIAHNOVQVPLQRLRIVRGTLFEDNYALAVLDNGD 121
DB 73 ITTYVHNRLDTFLKTIQEVAGVLIALLNVDVPIENLOIRGNVLYNSFALLVLSNH 132
QY 122 PLNNTPTVAGSPGGLRELQRLSFTLILKGVLIQIRNPOLCYQDTILMKDIFHKNNQAL 181
DB 133 -MNKTQ-----GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWMDIIDSRL-PL 182
QY 182 TLID-TNRSRACHPCSPCKGSRGSESDCQSLTRTVACGGCA-RCKGPLPNNFTVS 239
DB 183 TVLDFASNLSCPKCHPNCETDHCWGAGEQNCQTLTKVCAQQCSGRCKGVPSD----- 237
QY 240 FWLRVPKVSASHLE-----KHSDCLACLHFNHSGICELHCPALVYNTDTFESMNP 291
DB 238 -----CCHNQAAGCTGPRESDCLACRFRDATTCKTCCPLVLYNPTYQMDVNP 288
QY 292 EGRYTFGASCVTACPYNYLSTDVSGCTLVCPILHNOEVTAEADGTQRCCKSPCARVCYGL 351
DB 289 EGKYSFGATVCRECPHNVYWDHGSVCRSCNTDITYEV-EENGVRCKCKDGLCKVCNGI 347
QY 352 GMEHLREVRVATSANIQEFAGCKKIFGSLAFELPESFDGDPASNTAPLOEQVLPETLEE 411

```

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Plowman, G.D.; Culicou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epiderm

A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <P10>
A;Cross-references: GB:J07868; NID:G337359; PIDN:AAB59446.1; PID:G337360
A;Note: sequence extracted from NCBI backbone (NCBIP:126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;714-732/Region: protein kinase ATP-binding motif

Query Match 42.6%; Score 2905.5; DB 2; Length 1308;
Best Local Similarity 44.7%; Pred. No. 2.4e-117;
Matches 608; Conservative 181; Mismatches 376; Indels 195; Gaps 31;

QY 9 WGLLLALLPGAA-----STOVCTGTDMLKRLPASPETHLMDLRLHYOGCQVQVQGNLELT 64
DB 8 WVVSLVAAGTVQPSDSQSVCACTENKLSLSLEQQYRALRYKYYENCEVVMGNLEITS 67
QY 65 LPTNASLSFLQDIOEQVGYVLIHNOVRQVPLORLRIVRGTLQPLFEDNYVALAVLNGDPLN 124
DB 68 IEHRDLFSRVSREVTVGYVLVALNQRYLPLENLRIIRGTLYEDRYALAIFLNRYKDG 127
QY 125 NTPVTGASPGGLREQLRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQALTLI 184
DB 128 NF-----GLQELGLKNLTEILNGVGVYDQNKFLCYADTIHQDQIVRNWPNSLTLV 178
QY 185 DTNRSRACHPCSPMKGSRGWGSESDCOSLTRVTCAGGC-ARCKGPLPNNFTVSWLR 243
DB 179 STNGSSGGRCHKSGCTG-RWGPTEHNCOTLTRVCAEQDGRCYGY-----VS----- 227
QY 244 VPKVSASHLE-----KHSDDLACLFHNSHIGELHCPALVTYNTDTFESMNPBG 295
DB 228 ----DCCHRCAGCGSPKDTDFACMNFDSGACVTCQPTFYNYNTTTOLEHNFN 283
QY 296 TFGASCVTACPYNYLTDVSGSCTLVCPHNOEVTAEQDTCORCEKSKPCARVCYVGLQMEH 355
DB 284 TYGAFVCVKCPHNFV-VDSSSVCRAVCPSSRMEV-EENGIRKMKCPCTDICPKACDGI 341
QY 356 LREVRVTSANIOEFACGKIFGLSLAPESFGDPPASNTAPLOPEOLQVPELLEITGY 415
DB 342 LMSAQTVDSSNIQKFNCTKINGNLIFLVTGIHGDYPNATEADPEKLNFRVREITGF 401
QY 416 LYSAMPDLSLDSVFQNLQVIRGRIHLNGAYSILTLQGLGTSWLGSLRSLRELASGLAIH 475
DB 402 LNIQSMPPNMTDSVFSNLVTICGRVLYSGLSLILKQOGITSLOFQSLKEISGNIIYT 461
QY 476 HNTLCVHTVPDQLPRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCV 535
DB 462 DNSNLCTVHTINWTLTSTINQIRIVIRDRKAENCTAEGVMVNHLCSSDGCWGPGPQCL 521
QY 536 NCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECQ-ONGSVTCFGEADQCVACA 594
DB 522 SCRRFSRGRICISCNLDYGEFEFENGSI-CVECDPQCEKMEGDLTCHGPGDPNCTKCS 581
QY 595 HYKDPPEFCVACRSPGVKPDLSYMPIWKFPPDEEGACQPCINCTHSCVLDLDDKGC----- 648
DB 582 HFKXGPNVCEKCPDLQGANSF--IFKYADPDRECHPCNCTQCGNPTSHDCIYYPT 639
QY 649 -----PAEQRASPLTISVAVV-GILLVVGVVGVGILIKRQOKIRKYTMRLLOTEL 702
DB 640 GHSTLPQHAR-TPL--IAAGVIGGLFIVIVGLTFVAVYVRKSGIK-KKRALRRFL-ETEL 694
QY 703 VEPLTPSGAMPNOAMRIKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLR 762
DB 695 VEPLTPSGTAPNOAQRILKETELRKVKVLSGAFGVYKGIWIPDEGETKIPVAIKILN 754
QY 763 ENTSPKANKEILDEAVYVMAVGSPYSRLIGLICUTSTVQLVQTLMPYGLCLLDHVENRGR 822
DB 755 ETTGPKANVEFMDEALIMASMDPHLVRLVGLVCLSPITQLVQLMPHGCLLEYVHEKDN 814
QY 823 LGSQDILLNWCQIAKGMVLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETE 882

DB 815 IGSQDLLNWCQIAKGMVLEERRLVHRDLAARNVLVKSPNHVKITDFGLARLLDEGDEKE 874
QY 883 YHADGGKVPKIMMALESILRRRTHQSDVWSYGVYVWELMTFGAKPKPDGIPAREIPDLLE 942
DB 875 YNADGGKMPKIMMALECIHYRKETHQSDVWSYGVYVWELMTFGGKPYDGIPTREIPDLLE 934
QY 943 KGERLPOPPCTTIDVYIMVYKCMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEQ-LG 1001
DB 935 KGERLPOPPCTTIDVYIMVYKCMIDADSRPKFKELAAEFSESRMARDPQRYLVLIQDDRMK 994
QY 1002 PASPLDSTFYRSLLEDDDDMGDLVDAAEYLVPOQGFCCPDPAAGAGMVHHRHRSSTRSG 1061
DB 995 LPSNDKFPQNLLEDEEDLMDMAEYLVQ-QAFNIPPP-----IYTSRARDISNRS- 1046
QY 1062 GDLTGLLEPSEEEAPRS-----PLAP-SEGAGSDVP 1092
DB 1047 ----EIGHSPPPAYTPMSGNQFYVYRQDGAEGVSVYPYRAPTSTIPEAPVAQATAEIF 1102
QY 1093 DGDLGMAAGLQSLPHTDPSPLQRYSEDTPTVPLPS-----ETDGYVAPLTCSPQPEY 1145
DB 1103 DSCCNGTLRKVPAPHVQESSTQRYSDADPTVPAPERSPRGELDEEGYMTMRDKPKQEY 1162
QY 1146 VNQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAGGAVENPEYLTP 1205
DB 1163 LNPVE-----ENPFVSRR-----KNGDLQ-----ALDNPVEYHNA 1191
QY 1206 QCGAAAPQHPHPPA-----FSPAFDNLVYWDQ 1231
DB 1192 SNG-----PFAKAEDEYVNEPLYLNTFANTLGKAEYLNILSMPEKAKKAFDNPVYWNH 1245
QY 1232 DPPERGA--PPSTFKGTPT-----AENPEYL 1255
DB 1246 SLPRSTLQHPDYLOEYSTYFYKQNGRIRPVAENPEYL 1285

RESULT 8

S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transform
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Mittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Robe
Nature 341, 415-421, 1999
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <W1>
A;Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291
R;Adam, D.; Maeueller, W.; Scharf, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphoph
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285
C;Genetics:
A;Gene: mrk
A;Map position: Y
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; t
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif
Query Match 38.1%; Score 2602; DB 1; Length 1166;
Best Local Similarity 44.5%; Pred. No. 2.2e-104;


```
Db 224 NQCCHDEACGCGSPQDTCDFACRHFNDGACVPRCPQLVYNNKLTQLEPNPHTKYQG 283
Qy 299 ASCVTACPNYLSLTDVSGSCTLVCPHMQEVTAEGTQRCCKSPCARVCVGLGMEHLRE 358
Db 284 GVCVASCAPHNFV-VDTQSCVRACPPDRMEVD-KNGLKNCEPCGGLCPKACEGTSG--SR 339
Qy 359 VRVTSANIOBFACKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEEITGYLYI 418
Db 340 FQTVDSNIDCFVNTCKILGNLDFLTGLNDPWHKIPALDPEKLNVPRTVREITGYLNI 399
Qy 419 SAWPDSLPLSVPQNQLVIRGRIILHNGAYS-LTQLGLGISWGLRSLRSLRELGLALHNN 477
Db 400 QSWPPHMHNFVSNLTTIGRSLYNRGFSLLIMKNLNVTSLSGRSLKEISAGRIYISAN 459
Qy 478 THLCFVHTVPWDQLFRPHQALLHTA-NRPDECVGEGGLACHQICARGCHGPGPTQCVN 536
Db 460 RQLCYHHSNLNWKVLRGTERTLDIKHNRPRDCVAEGKVDPLCSSGGCGKPGPGQCLS 519
Qy 537 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCVACAHY 596
Db 520 CRNYSRGVCVTHCNFLNGEPREFAHEAECFCHPECPQMGSTATCNGSGSDTCAQCAPH 579
Qy 597 KDPFCVACRSGVKPDLSYNMIWPKPDEGACQPCPNCTHSCVDLDDKCPAEQRA-- 654
Db 580 RDGPHCVSSCPHGVLG--AKGPIYKYPDVQNECRPCHENCQGGCKGPELQDCLGQTLVLI 637
Qy 655 --SPLTSIVSAVGLLVVGLVGVGILI KEROQKIR-KYTWBRLLOQTELVEPLTPSGA 711
Db 638 GKTHLTWALTVIAG--LVVIFMLGGTFLYWRGRIQNKWARRRYLERGESIEPLDPS-E 694
Qy 712 MPNQQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPRANK 771
Db 695 KANKVLARIKPETELKLVKLVGSGVGVGVHGVWIPGESIKIPVICKVIEDKSGRQSFQ 754
Qy 772 EILDEAYMAGVSGPYVSRLLGICLTSTVQLVQLMPYGCILLDHVRENRLGSGQDLLNW 831
Db 755 AVTDHMLAIGSLDHAHIVRLGLCPGSSLQVLTQVPLGSLLDHVHQHRAIGALGPQLLW 814
Qy 832 CMQJAKGMSYLEDVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 891
Db 815 GVQJAKGMYLLEEHGMHNRNLAARNVLKSPQOVADFVADLLPPDKQLLYSEAKTP 874
Qy 892 IKWMALESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 951
Db 875 IKWMALESIHFGKYTHQSDVMSYGVTVWELMTFGAEPYAGRLAEVPPDLLEKGERLA 934
Qy 952 ICTTDVYMWKCMWIDSECRPRPRELVSEFSRWARDPQRVFIQNEIDLGA---SPLDS 1008
Db 935 ICTTDVYMWKCMWIDENIRPTEKELANEFTRWARDPPRYLVIKRES-GPGIAPGPEPH 993
Qy 1009 TFYRSLLDDMDGLVDAEEVLPQOGFFCPDPAPGAGGMVHRRHSSSTRSGGDLTLG 1068
Db 994 GLTNKULEEVELEBELDLDLDAEED-----NLATTLGASLSLP 1034
Qy 1069 LEP-SEEAAPRSLAPSEAGSDVFDGLGMAAGLQSLPETHD-PSPLQRYSDPTVPL 1126
Db 1035 VGTLLNRPRGSQSLSPSSGY-MPMNQGNLGSQESAVSGSSERCPRVPSLH-----PM 1087
Qy 1127 P-----SETDGYA-----PLTCSPOPE-----YVNPQDVRPQPPSPRE 1160
Db 1088 PRGCLASESGHYTGSEAELOEKVSMCRSRSRSPRPGDSAYHSQRHSLLPVTPVPLS 1147
Qy 1161 GP-----LPAARPAATLERAKTLLSP-GKNGV-----KDVFAFGAVENPEY 1202
Db 1148 PPGLEEDVNGVMPDTHKTKTPSRECTLSSVGLSSVLGTEEBED-----EVEY 1199
Qy 1203 LTPQGAAPQHPPPAFSPAFDNLYYND-----QDPPERGAPPTFKGTPTAE 1250
Db 1200 MNRRRRSP-PHPPRPSLEELGYEYMDVGSLSLSTQSCPLHPVPIMPTAGTTTPE 1258
Qy 1251 NPEYL 1255
Db 1259 DYEYM 1263
```

RESULT 10

JC4387

epidermal growth factor receptor homolog precursor - rat

N:Alternate names: ErbB3 protein; HER3 protein

C:Species: Rattus norvegicus (Norway rat)

C:Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998

C:Accession: JC4387

R:Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.

Gene 165, 279-284, 1995

A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protei

A:Reference number: JC4387; MUID:96096535; PMID:8522190

A:Accession: JC4387

A:Molecule type: mRNA

A:Residues: 1-1339 <HEL>

A:Cross-references: GB:029339; NID:g915309; PID:g915390

A:Experimental source: liver

A:Note: The authors translated the codon AAC for residue 369 as Thr and GTT for res

C:Comment: This protein is a functional heregulin receptor that transduces signals t

C:Genetics:

A:Gene: ErbB3

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hor

C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>

F:640-659/Domain: transmembrane #status predicted <TMM>

F:705-970/Domain: protein kinase homology <KIN>

F:713-721/Region: protein kinase ATP-binding motif

F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)

Query Match

Best Local Similarity 32.9%; Score 2248.5; DB 2; Length 1339;

Matches 516; Conservative 168; Mismatches 436; Indels 177; Gaps 36;

```
Qy 3 LAALCRWGLLIALLPPGAA---STQVCTGTDMKLRLLPASPEHLDMRLHLYOGCOVQGN 59
Db 7 LOVLG---FLLSLARGSEMNSOAVCPGLTNGLSVTGDADNOVQTLKLYEKEVWGN 62
Qy 60 LELTYLPTNASLSFLQDIOEVQGYLYIAHNOVROVLOPLRIVRGTLQTFEDNYALAVLDN 119
Db 63 LEIVLTGHADLSFLQWIREVYAVLVAMNEFSLVPLPNLRVVRGTQVYDGKFAIVM-- 120
Qy 120 GDPLNNTPVTGASPGGLRELOLSLTELKGGVLIQBNPOLCYQDTILMKWDIIPHKNQL 179
Db 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLCMDTIDWRDIVRVR--- 170
Qy 180 ALTLLIDNRSRACHPCSPMKSGRCWGESSEDCOSLTRTVACAGG-ARCKGPLPNNFTV 238
Db 171 GABIVVKNNGANGCPCHVECKG-RCWGGPDDCQILTKTICAPQCNGRCFGNP----- 223
Qy 239 SFWLVRPKVASHLE-----KHSCLACLHFNHSGICELHCPALVYNTDTPESHNP 290
Db 224 -----NQCCHDEACGCGSPQDTCDFACRHFNDGACVPRCPQLVYNNKLTQLEPN 275
Qy 291 PEGRYTFGASCVTACPNYLSLTDVSGSCTLVCPHMQEVTAEGTQRCCKSPCARVCVCG 350
Db 276 PHTKYQGVGVASCAPHNFV-VDTQFCVRACPPDRMEVD-KHGLKNCEPCGGLCPKACBG 333
Qy 351 LGMEHLREVRAVTSANIOBFACKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLE 410
Db 334 TSGG--SRYQTVDSNIDGFCVNTCKILGNLDFLTGLNDPWHKIPALDPEKLNVPRTV 391
Qy 411 EITGYLISAWPDSLPLSVPQNQLVIRGRIILHNGAYS-LTQLGLGISWGLRSLRSLREL 469
Db 392 EITGYLISAWPDMHNFVSNLTTIGRSLYNRGFSLLIMKNLNVTSLSGRSLKEISA 451
Qy 470 GLALIHNTHLCPVHTVPWDQLFRPHQALLHTA-NRPDECVGEGGLACHQICARGCHWG 528
Db 452 GRVYSANQQLCYCHHSNLNWKVLRGTERTLDIKHNRPRDCVAEGKVDPLCSSGGCG 511
Qy 529 PGPTQCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPAD 588
```

Db 512 PAPGQCLSCRNYSREGVCVTHCNFLOQEPREFVHEAQCPSCHPECLPMGECTSTYNGSGSD 571
QY 589 QCVACAHYKDPFCVACRPSGVKPDLSYMPKWPDEGACQPCPINCTHSC--VDLDDK 646
Db 572 ACARCAHFRDGHCVNSCPHGLG--AKGPIYKYPDAQNECRCHENCCTQCGNPELQDC 629
QY 647 GCPAEQRASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIR-KYTMRLLOETELVER 705
Db 630 LQAEVLMKPHLVIAVTVG--LAVILMLGSGFLYWRGRRQNKAMRYLGERGESIEP 687
QY 706 LTPSGAMPNQAOQRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVURENT 765
Db 688 LDPS-EKANKVLARIFKETELRKVLKVLGSGVGTGKHGIWIPGESIKIPVCIKVIEDKS 746
QY 766 SPKANKEILDEAVVMAGVSPYVSRLLGICLTSTVOLVTQLMYPGCLLDHVRNRRGLS 825
Db 747 GRQSOAVTDHMLAVGLDHAHIVRLGLCPGSSQLQVLTQYLPGLSLLDHVKQHRETLGP 806
QY 826 QDLLNMCIOAKGMSYLEVDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHA 885
Db 807 QULLNGVQIAGMYLYLESHVHRDLAARNVLMKSPQVQVADFGVADLLPDDKQLLH 866
QY 886 DGGKVPKWMALLESILRRRFTHQSDVMSYGVTVWELMTFCAPYDGIPIAREIPDLLEKGE 945
Db 867 SEAKTPIKWMALLESIFGKYTHQSDVMSYGVTVWELMTFGAPYAGLRRLAEIPDLLEKGE 926
QY 946 RLPOPPICITDVMIMVKWMDIDSECRPRFRELVSFSRMARPORFVVIQNEDLGPASP 1005
Db 927 RLQAQOICITDVMIMVKWMDIDENIRPTFKELANEFTRMARDPPRYLVKRAAS-GPGTP 985
QY 1006 LOSTFYRSILLEDDMGDLVDAEYLVPOQGFPCDPAPGAGGVHHRHSSSTRSGGDL 1065
Db 986 --PAEAPSVLTTKEL-----QEALEPEL-----DL 1009
QY 1066 TLGLEPSE-----EAPRSPLAPSEG-----AGSDVDFCDLG 1097
Db 1010 DLDLEAEGLATSLGALSLLPTGLTRPRGSQLSPSSGYMPMQSSIGEACLDSAVL 1069
QY 1098 MGAAGLQSLPHDPSPLQRYSEDTPVLPSETDGVV-----APL-----TC-----SP 1141
Db 1070 GGREQFSRPLSLH-PIPRGR-----PASESEGHVGTGSEALQEKVSVCRSRSRSP 1121
QY 1142 QPE---VYNQPDVRPQPSPREGP-----LPAARPAGATLERAKTLSP-GKNG 1185
Db 1122 RRGDSAYHSQRHSLTPTPLSPGLEEDNGVYMPDTHLRGASSSREGTLSSVGLSS 1181
QY 1186 VV-----KDVFAFGAVENPEYLTPOGGAAPQPHPP 1216
Db 1182 VLGTEEEDE-----EEYEYMNKRKRGSP-PRPP 1209
RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nielsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <NIL>
R:Cross-references: GB:M10066; GB:M13881; NID:g211749; PID:AAA48763.1; PID:g211750
A>Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F:1-6/Product: gag protein (fragment) #status predicted <GAG>

F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F:194-459/Domains: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted

Query Match 25.9%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 8.1e-69;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

QY 584 GPEADOCVACAHYKDPFCVACRPSGVKPDLSYMPKWPDEGACQPCPINCTHSCVDL 643
Db 60 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VKVDADANAVCOLCHPNCNTRCKGKP 116
QY 644 DDKGCPAEQRASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTMRLLOETEL 702
Db 117 GLEGCP---NGSKTSPISAGVVGGLLCLVVGIGLGLRRR-HIVKRTLRLLQOREL 172
QY 703 VPLTPSGAMPNQAOQRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLR 762
Db 173 VEPLTPSGEAPNQAHILKETETEFKKVKVLGSGAFGTVYKGLWIPEGEKVIPVAIKELR 232
QY 763 ENTSPKANKEILDEAVVMAGVSPYVSRLLGICLTSTVOLVTQLMYPGCLLDHVRNRR 822
Db 233 EATSPKANKEILDEAVVMASVDNPHVCRLLGICLTSTVOLVTQLMYPGCLLDVIREHKN 292
QY 823 LGSQDLLNMCIOAKGMSYLEVDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
Db 293 IGSQYLLNMCVQIAGMYLYLESHVHRDLAARNVLMKSPQVQVADFGVADLLPDDKGADEKE 352
QY 883 YHADGKVPKWMALLESILRRRFTHQSDVMSYGVTVWELMTFCAPYDGIPIAREIPDLLE 942
Db 353 YHAEGKVPKWMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLE 412
QY 943 KGERLPOPPICITDVMIMVKWMDIDSECRPRFRELVSFSRMARPORFVVIQNEDLG 1001
Db 413 KGERLPOPPICITDVMIMVKWMDIDADSRPKFELIAEFSKMARPPRYLVITQDGRMH 472
QY 1002 PASPLDSTFYRSILLEDDMGDLVDAEYLVPOQGFPCDPAPGAGGVHHRHSSSTRSG 1061
Db 473 LPSPTDSKYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 513
QY 1062 GGDLTGLLEPSEEEAPRSPL-----APSEAGSDVDFDGLMGAAKGLQSLPHDPSPLQ 1116
Db 514 -----SRTPLSSLSATSNNSATNCID-----RNGQGHVREDSFVQ 550
QY 1117 RYSEDPVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPRGPLPAARPAGATLE 1174
Db 551 RYSSDPTGNFLEESIDDGFL-----PAPEYVYVQ--LMPKKPS----- 585
QY 1175 RAKTLSPGKNGVVKVDF-----AFGAVENPEYLTPOGGAAPQPHPPAF 1219
Db 586 ----TAMVQVQIYNNISLTAISKLPWDSRYQNSHSTAVDNPEYL-----NTNOSPLA 633
QY 1220 SPAFONLYYWDQ-----DPPE-----RGAPSTFKGTPTAENPEYLGIDVP 1260
Db 634 KTVFESSPWVIQSGNHQINLDNPDYQDDFLPNETKPNGLLKYPAAENPEYLRVAAP 689

RESULT 12

TVUWH

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fa
A:Reference number: A00644; MUID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678

R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, Science 224, 1456-1459, 1994
A;Title: Sequencing the erba gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022

A;Molecule type: DNA
A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A;Cross-references: GB:K02066
C;Genetics:

A;Gene: erbB

C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

Query Match 24.9%; Score 1703; DB 1; Length 604;

Best Local Similarity 52.2%; Pred. No. 3.7e-66;

Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

QY 593 CAHYKDPFVCVACPCSGKPDLSYMPIWKPPDEEGACQPCPINTHSCVLDLDDKGCPEAQ 652

DB 3 CAHFDGPHCVKACPAVLGENDTL-VRYKADANAVCOLCHPNCTRCCKPGLEGCP--- 58

QY 653 RASPLTSIVSAVV-GILLVVVLGVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGA 711

DB 59 NGSKTPSIAAGVVGGLLCLVWGLGIGLYLRR-HIVKRTLRLLQERELVEPLTPSGE 117

QY 712 MPNOAOMRLKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 771

DB 118 APNOAHLRIKETEFKKVKVLGSGAGFTIYKGLWIPEGEKVIIPVAIKELREATSKANK 177

QY 772 EILDEAYVMAGVSPVYSRLGICLTSTVOLVTLQMPYGLLDHVRNRLGSGQDLLNW 831

DB 178 EILDEAYVMASVDNPHVCRLLGICLTSTVOLITQLMPYGLLDYIYREHKONISQYLLNW 237

QY 832 CMQIAGKMSYLEDLVLRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEYHADGGKVP 891

DB 238 CVQIAGKMYLEBRLVLRDLAARNVLVKTPOHVKITDFGLAKLGLGADKEHYHAEGGKVP 297

QY 892 IKWMALESILRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPP 951

DB 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTGSKPYDGIPIAREIPDLLEKGERLPQPP 357

QY 952 ICTIDVYIMVKWMDSECRPRFELVSEFSRMARDPQRFVVIQ-NEDLGASPLDSTF 1010

DB 358 ICTIDVYIMVKWMDIADSRPFRELIAEFSKWARDPPRYLVLIQDGRMHLSPDTSKF 417

QY 1011 YRSLLDDDDMDLVDAEYLVPOQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLE 1070

DB 418 YRTLMBEEDMEDIVDAEYLVPHQGF-----NSPST----- 449

QY 1071 PSEEEAPRSP-----APSEGAGSDVFDGLGMAAKGLQSLTHDPSPLQRYSEOPTVP 1125

DB 450 -----SRTPLLSLSLTSATSNATNCID-----RNGQHPVREDSFVQRYSSDPTGN 495

QY 1126 LPSET--DGYPVAPLTCSPQPEYVQDVRPQPPSPREGPLPAARAGATLERAKTISPGK 1183

DB 496 FLEESIDDGFL-----PAPEYVQ--LMPKPKSTAM----- 524

QY 1184 NGVVKDVFAF-----GGAVENPEYLTPOGGAAPQPHPPAFSPAFD 1224

DB 525 --VQNGIYNFISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPKATVFE 574

QY 1225 NLYVWDQDPPERCAPPSTFKGTTAENPEY 1254

DB 575 SSPYWIQSGNHQ-----INLONPDY 594

RESULT 13

S35745

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus

C;Species: avian erythroblastosis virus

C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C;Accession: S35745

R;Vennstrom, B. EMBL Data Library, March 1993

Submitted to the EMBL

A;Reference number: S35743

A;Accession: S35745

A;Molecule type: DNA

A;Residues: 1-544 <VEN>

A;Cross-references: EMBL:X12707

C;Genetics:

A;Gene: erbB

C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

Query Match 24.1%; Score 1647; DB 2; Length 544;

Best Local Similarity 54.9%; Pred. No. 8.3e-64;

Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 584 GPADOCVACAHYKDPFVCVACPCSGVKPDLSYMPIWKPPDEEGACQPCPINTHSCVDL 643

DB 1 GP--DHCMKCAHPIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHPNCTRCCKGP 57

QY 644 DDGKCPAEQASPLTSIVSAVV-GILLVVVLGVFGILIKRQOKIRKYTMRLLOETEL 702

DB 58 GLEGCP---NGSKTPSIAAGVVGGLLCLVWGLGIGLYLRR-HIVKRTLRLLQEREL 113

QY 703 VEPLTPSGAMPNOAOMRLKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLR 762

DB 114 VEPLTPSGEAPNQAHLRIKETEFKKVKVLGSGAGFTVYKGLWIPEGEKVIIPVAIKELR 173

QY 763 ENTSPKANKEILDEAYVMAGVSPVYSRLGICLTSTVOLVTLQMPYGLLDHVRNRR 822

DB 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLITQLMPYGLLDYIREHKON 233

QY 823 LGSODLLNMCQIAGKMSYLEDLVLRDLAARNVLVKSPPNHVKITDFGLARLLDIDETE 882

DB 234 IGQYLLNWCQIAGKMYLEBRLVLRDLAARNVLVKTPOHVKITDFGLAKLGLGADKE 293

QY 883 YHADGGKVPKWMALLESILRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLE 942

DB 294 YHAEGGKVPKWMALLESILHRIYTHQSDVMSYGVTVWELMTGSKPYDGIPIAREIPDLLE 353

QY 943 KGERLPQDPICITIDVYIMVKWMDSECRPRFELVSEFSRMARDPQRFVVIQ-NEDLG 1001

DB 354 KGERLPQDPICITIDVYIMVKWMDIADSRPFRELIAEFSKWARDPPRYLVLIQDGRMH 413

QY 1002 PASPLDSTFVRSLLDDDDMDLVDAEYLVPOQGFPCPDPAFCAGGMVHRRSSSTRSG 1061

DB 414 LPSPTDSKPYRTLMBEEDMEDIVDAEYLVPHQGF-----NSPST--- 454

QY 1062 GGLTLGLEFSEEEAPRSP-----APSEGAGSDVFDGLGMAAKGLQSLTHDPSPLQ 1116

DB 455 -----SRTPLLSLSLTSATSNATNCIDRNGG-----H----- 481

QY 1117 RYSEDTVPVLPSETDGYVAPLTCSPQPEYVQDVRPQPPSPREGPLPAARAGAT-LER 1175

DB 482 -----PVREDGFL-----PAPEYVQ--LMPKPKSTAMVQNIYISLTAISK 523

QY 1176 AKTLPKGVKGVKDVFAFGGAVENPEYL 1203

DB 524 LPIDSRYN-----SHSTAVDNPEYL 544

RESULT 14

S00727

kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis viru

C;Species: avian erythroblastosis virus

C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997

C;Accession: S00727

R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.

A:Accession: B44776
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-540 <BRU>
A:Cross-references: GB:X52211
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: Atp; phosphotransferase; tyrosine-specific protein kinase
F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif

Query Match 23.8%; Score 1623; DB 2; Length 540;
Best Local Similarity 54.9%; Pred. No. 8 8e-63;
Matches 340; Conservative 69; Mismatches 120; Indels 90; Gaps 14;

Qy 593 CAHYKDPPFCVACRPGVKPDLSYMPFWPDPDEGACQPCPNCTHSCVDLDDKGCPAEQ 652
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 3 CAHFIDGPHCVKACPAGVGENDTL-VMKYADANAVCLCHPNCTRGGKPGLEGCP-- 58

Qy 653 RASPLTSIVSAVV-GILLVVVLGVVFGLIKRRQQIKRYKTHRRLLQETELVEPLTPSGA 711
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 59 NGSKTSPSIAAGVGGGLLVGLIGLGLYLRRI-HTRKRITLRLQLQERLEVEPLTPSGE 117

Qy 712 MPNQAMRIILKETELRKVKVLGSCAFGTGYKGIWIPDGENVKIPVAIKVRENTSPANK 771
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 118 APQAHRLIILKETEFKKVKVLGSCAFGTGYKGLWIPEGEKVITIPVAIKEUREATSPANK 177

Qy 772 EILDVAYVMAGVGSPPVSRLLGLICLTSTVOLVTQLMPYGCLLDHVRENRGSLSDLLNN 831
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 178 EILDVAYVMASVDNPHVCRLGLICLTSTVOLITQLMPYGCLLDYIREHKDNIGSQVLLNW 237

Qy 832 CMQIAKMSYLEVDRLVHRDLAARNVLVKSFNHVKITDFGLARLLDIDETEHADCGKVP 891
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 238 CVQIAKGMNYLEERHMVHRDLAARNVLVKTPOHKVITDFGLAKQLGADEKEYHAEGKVP 297

Qy 892 IKMWALLESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKERLPQPP 951
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 298 IKMWALLESILHRYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLEKERLPQPP 357

Qy 952 ICTIDVYIMVKCMWIDSECRPRFRELVSEFSRMARDPQRFVVIQ-NEDLGPA SPLDSTF 1010
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 358 ICTIDVYIMVKCMWSGADSRPKRELIAEFCKMARDPPRYLVIOGDERMHLPSTDSKF 417

Qy 1011 YRSILLEDDMGDLVDAAEYLVPQGFCDPDAPGAGMVHRRSSSTRSGGDLTGLGE 1070
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 418 YRTLMEEDMEDIVDAEYLVPHQGFF-----NSPST----- 449

Qy 1071 PSEEAPRPL-----APSEGAGSDVFDGDLNGAAKGLQS LPTHDPSPLORYSEDPTVP 1125
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 450 -----SRTPLLSSLSATSNNSATNCIDRNG-----H----- 476

Qy 1126 LPSETDGYVAPLTCSPQSYVNQPDPVRQPPSPREGPLPAARPAGAT-LERAKTLSPGN 1184
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 477 -PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNQIYNVISLTAISKLPMDRYQN 527

Qy 1185 GVVKQVFAFGGAVENPEYL 1203
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 528 -----SHSTAVDNPEYL 539

Search completed: July 22, 2003, 09:08:59
Job time : 31.0157 secs

Search completed: July 22, 2003, 09:08:59
Job time : 31.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-210-224-14

Perfect score: 6827

Sequence: 1 MELAALCRWGLLLALLPPCA.....TFKGTPTAENPEYLGLDVDPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|--------------|--------------------|
| 1 | 6684 | 97.9 | 1255 | 1 ERB2 HUMAN | P04626 homo sapien |
| 2 | 5872 | 86.0 | 1257 | 1 ERB2 RAT | P06494 rattus norv |
| 3 | 5870.5 | 86.0 | 1254 | 1 ERB2 MESAU | Q60553 mesocricetu |
| 4 | 3052 | 44.7 | 1210 | 1 EGFR_HUMAN | P00533 homo sapien |
| 5 | 3031 | 44.4 | 1210 | 1 EGFR_MOUSE | Q01279 mus musculu |
| 6 | 2905.5 | 42.6 | 1308 | 1 ERB4_HUMAN | Q15303 homo sapien |
| 7 | 2886 | 42.3 | 1308 | 1 ERB4 RAT | Q62956 rattus norv |
| 8 | 2618.5 | 38.4 | 1167 | 1 XMRK_XIPMA | P13388 xiphophorus |
| 9 | 2342.5 | 34.3 | 1342 | 1 ERB3_HUMAN | P21860 homo sapien |
| 10 | 2271.5 | 33.3 | 1339 | 1 ERB3 RAT | Q62799 rattus norv |
| 11 | 1883 | 27.6 | 1426 | 1 EGFR_DROME | P04412 drosophila |
| 12 | 1749.5 | 25.6 | 634 | 1 ERBB_ALV | P00534 avian leuko |
| 13 | 1703 | 24.9 | 604 | 1 ERBB_AVIER | P00535 avian eryth |
| 14 | 1630 | 23.9 | 540 | 1 ERBB_AVIEU | P11273 avian eryth |
| 15 | 1510 | 22.1 | 703 | 1 EGFR_CHICK | P13387 gallus gall |
| 16 | 1351 | 18.3 | 1323 | 1 LTR3_CABEL | P24348 caenorhabdi |
| 17 | 1142.5 | 16.7 | 245 | 1 ERB2_MOUSE | P70424 mus musculu |
| 18 | 678 | 9.9 | 1382 | 1 INSR_HUMAN | P06213 homo sapien |
| 19 | 671 | 9.8 | 1363 | 1 INSR_BRALA | Q02466 brachiosto |
| 20 | 671 | 9.8 | 1383 | 1 INSR_RAT | P15127 rattus norv |
| 21 | 670.5 | 9.8 | 1372 | 1 INSR_MOUSE | P15127 rattus norv |
| 22 | 662 | 9.7 | 1297 | 1 IRR_HUMAN | P15208 mus musculu |
| 23 | 661 | 9.7 | 1607 | 1 MIPR_LYMST | P14616 homo sapien |
| 24 | 659 | 9.7 | 1300 | 1 IRR_MOUSE | Q25410 lymnaea sta |
| 25 | 656.5 | 9.6 | 1300 | 1 IRR_CAVPO | Q9wt14 mus musculu |
| 26 | 643 | 9.4 | 1477 | 1 HTK7_HYDAT | P14617 cavia porce |
| 27 | 604 | 8.8 | 1367 | 1 IGIR_HUMAN | Q25197 hydra atten |
| 28 | 597 | 8.7 | 1373 | 1 IGIR_MOUSE | P08069 homo sapien |
| 29 | 593.5 | 8.7 | 1370 | 1 IGIR_RAT | Q60751 mus musculu |
| 30 | 591 | 8.7 | 987 | 1 EPB4_HUMAN | P24062 rattus norv |
| 31 | 589.5 | 8.6 | 977 | 1 EPB2_MOUSE | Q03145 mus musculu |
| 32 | 588 | 8.6 | 1114 | 1 RET_HUMAN | P07949 homo sapien |
| 33 | 588 | 8.6 | 1390 | 1 INSR_AEDAE | Q93105 aedes aegyp |

RESULT 1

| ID | ERB2_HUMAN | STANDARD; | PRT; | 1255 AA. |
|----|--|-----------|------|----------|
| AC | P04626; | | | |
| DT | 13-AUG-1987 (Rel. 05, Created) | | | |
| DT | 13-AUG-1987 (Rel. 05, Last sequence update) | | | |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update) | | | |
| DE | Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112) | | | |
| DE | (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell | | | |
| DE | surface receptor HER2) (MLN 19). | | | |
| GN | ERBB2 OR HER2 OR NGL OR NEU. | | | |
| OS | Homo sapiens (Human) | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | |
| OX | NCBI_TaxID=9606; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=86118663; PubMed=3003577; | | | |
| RA | Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N., | | | |
| RA | Saito T., Toyoshima K.; | | | |
| RT | "Similarity of protein encoded by the human c-erb-B-2 gene to | | | |
| RT | epidermal growth factor receptor."; | | | |
| RL | Nature 319:230-234(1986). | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=86070181; PubMed=2999974; | | | |
| RA | Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A., | | | |
| RA | McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J., | | | |
| RA | Frankie U., Levinson A., Ullrich A.; | | | |
| RT | "Tyrosine kinase receptor with extensive homology to EGF receptor | | | |
| RT | shares chromosomal location with neu oncogene."; | | | |
| RL | Science 230:1132-1139(1985). | | | |
| RN | [3] | | | |
| RP | SEQUENCE OF 737-1031 FROM N.A. | | | |
| RX | MEDLINE=86016729; PubMed=2995967; | | | |
| RA | Semba K., Kamata N., Toyoshima K., Yamamoto T.; | | | |
| RT | "A v-erbB-related protooncogene, c-erbB-2, is distinct from the | | | |
| RT | c-erbB-1/epidermal growth factor-receptor gene and is amplified in a | | | |
| RT | human salivary gland adenocarcinoma."; | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985). | | | |
| RN | [4] | | | |
| RP | VARIANTS VAL-654 AND VAL-655. | | | |
| RX | MEDLINE=93194196; PubMed=8095488; | | | |
| RA | Ehsani A., Low J., Wallace R.B., Wu A.M.; | | | |
| RT | "Characterization of a new allele of the human ERBB2 gene by allele- | | | |
| RT | specific competition hybridization."; | | | |
| RL | Genomics 15:426-429(1993). | | | |
| CC | !- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, | | | |
| CC | ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A | | | |
| CC | POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF- | | | |
| CC | ALPHA AND AMPHIREGULIN. | | | |
| CC | !- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein | | | |
| CC | tyrosine phosphate. | | | |
| CC | !- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS | | | |
| CC | (POTENTIAL). | | | |
| CC | !- SUBCELLULAR LOCATION: Type I membrane protein. | | | |

ALIGNMENTS

| | | | | | |
|----|-------|-----|------|---|------------|
| 34 | 584.5 | 8.6 | 976 | 1 | EPB2_HUMAN |
| 35 | 583.5 | 8.5 | 984 | 1 | EPB1_CHICK |
| 36 | 583 | 8.5 | 987 | 1 | EPB4_MOUSE |
| 37 | 581 | 8.5 | 2146 | 1 | INSR_DROME |
| 38 | 573.5 | 8.4 | 1053 | 1 | PAK1_CHICK |
| 39 | 572.5 | 8.4 | 984 | 1 | EPB1_RAT |
| 40 | 569 | 8.3 | 1088 | 1 | PAK1_XENLA |
| 41 | 567 | 8.3 | 757 | 1 | HT16_HYDAT |
| 42 | 566.5 | 8.3 | 984 | 1 | EPB1_HUMAN |
| 43 | 564.5 | 8.3 | 985 | 1 | EPB4_XENLA |
| 44 | 564 | 8.3 | 902 | 1 | EPBB_XENLA |
| 45 | 563 | 8.2 | 1052 | 1 | PAK1_MOUSE |

| | |
|--------|-------------|
| P29317 | homo sapien |
| Q07494 | gallus gall |
| P54761 | mus musculu |
| P09208 | drosophila |
| Q00944 | gallus gall |
| P09759 | rattus norv |
| Q91738 | xenopus lae |
| P53356 | hydra atten |
| P54762 | homo sapien |
| Q91571 | homo sapien |
| Q91736 | xenopus lae |
| P34152 | mus musculu |


```
QY 713 PNOAQRILKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 772
Db 707 PNOAQRILKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 766
QY 773 ILDEAYMAGVGSFYVRLGICLCTSTVQLVLTOLMPYVGCCLLDHVRENRGLSGQDLLNWC 832
Db 767 ILDEAYMAGVGSFYVRLGICLCTSTVQLVLTOLMPYVGCCLLDHVRENRGLSGQDLLNWC 826
QY 833 MQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDIDETEHADGGKVP 892
Db 827 MQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDIDETEHADGGKVP 886
QY 893 KWALESILARRRTHOSDVMSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQPI 952
Db 887 KWALESILARRRTHOSDVMSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQPI 946
QY 953 CTIDVTMVMKWMIDSECRPRELVSFERNARDPQRFVVIQNEDELGPASPLDSTFYR 1012
Db 947 CTIDVTMVMKWMIDSECRPRELVSFERNARDPQRFVVIQNEDELGPASPLDSTFYR 1006
QY 1013 SLEDDDDMGDLVDAEYLYPQGGFCFDPDPAGAGMVRHRRSSSTRSGGDLTLGLEPS 1072
Db 1007 SLEDDDDMGDLVDAEYLYPQGGFCFDPDPAGAGMVRHRRSSSTRSGGDLTLGLEPS 1066
QY 1073 EEAAPRSPLAPSGAGSDVPDGLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDG 1132
Db 1067 EEAAPRSPLAPSGAGSDVPDGLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDG 1126
QY 1133 YVAPLTCSPQPEVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSCNKGWVDVFA 1192
Db 1127 YVAPLTCSPQPEVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSCNKGWVDVFA 1186
QY 1193 FGAIVENPEYLTQGGAAPOHPPPAFSPAFDNLTYWDQDPPPERGAPPSTFKGTPTAENP 1252
Db 1187 FGAIVENPEYLTQGGAAPOHPPPAFSPAFDNLTYWDQDPPPERGAPPSTFKGTPTAENP 1246
QY 1253 EYLGLDVPV 1261
Db 1247 EYLGLDVPV 1255

RESULT 2
ERB2_RAT
ID ERB2_RAT STANDARD; PRT; 1257 AA.
AC P06494;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE receptor-related protein).
GN ERB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=86118662; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The new oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
```

```
RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RT oncogenic and oncogenic forms of the neu protein.";
RL EMBO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC ENBL; X03362; CAA27059.1; ALT_INIT.
DR PIR; A24562; TVRTNU.
DR HSSP; P11362; LFQG.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR Pfam; PF02757; YLP_2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 677 POTENTIAL.
FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 369 CYS-RICH.
FT DOMAIN 473 646 CYS-RICH.
FT DOMAIN 722 989 PROTEIN KINASE.
FT NP_BIND 728 736 ATP (BY SIMILARITY).
FT BINDING 755 755 ATP (BY SIMILARITY).
FT ACT_SITE 847 847 BY SIMILARITY.
FT DISULFID 196 205 BY SIMILARITY.
FT DISULFID 200 213 BY SIMILARITY.
FT DISULFID 221 228 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 312 BY SIMILARITY.
FT DISULFID 316 332 BY SIMILARITY.
FT DISULFID 335 339 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
```

```
FT DISULFID 517 530 BY SIMILARITY.
FT FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.0%; Score 5872; DB 1; Length 1257;
Best Local Similarity 85.8%; Pred. No. 3.7e-311;
Matches 1090; Conservative 51; Mismatches 106; Indels 24; Gaps 4;

Qy 1 MELAALCRWGLLLALLPPGAASCTVCTGDMKRLPASPETHLDMLRHLRYGCGQVQGNL 60
Db 1 MELAAWCRWGLLLALLPPGIAGTCTGDMKRLPASPETHLDMLRHLRYGCGQVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIQEOGYVLIAHNQVROVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
Db 61 ELTYVPANASLSFLQDIQEOGYVLIAHNQVRLPLQRLRIVRGTOQLFEDNYALAVLDNR 120

Qy 121 DPLNNTTPVT-GASPGGLRELQRLSLTEILKGGVLIQIRNPOLCYODTILKWDIFHKNNOL 179
Db 121 DPQNVAASTPRTPEGLRELQRLSLTEILKGGVLIQIRNPOLCYQDWLWKDVRKNNOL 180

Qy 180 ALTLIDNRSRACHPCSPMKGSRGWESSEDCQSLTRTVCAAGCARCKGLPLFNNFTVS 239
Db 181 APVDIDNRSRACPPCAPACKDNHCWGESPEDCQILTGITCTGCARCKGLP----- 233

Qy 240 FWLRVPKVSASHLE-----KHSDDLCLHFNHSGICELHCPALVTYNTDTFESMNP 291
Db 234 -----TDCHEOQAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMNP 286

Qy 292 EGRYTFGASCVTCPYNYLSDVGSCTILVCPHNLQEVTAEDGTORCEKCKSPCARVCYGL 351
Db 287 EGRYTFGASCVTCPYNYLSTEVGSCTILVCPNNQEVTAEDGTORCEKCKSPCARVCYGL 346

Qy 352 GMEHLRVRVTSANIOEFACCKKIFGSLAFPLSPFDGDPASNTAPLOEQLOVFETLEE 411
Db 347 GMEHLRGARAITSDNVQEFDCCKKIFGSLAFPLSPFDGDPSSGIAPLRPEOLOVFETLEE 406

Qy 412 ITGYLYISAMPDSLDPILSVFONLQVIRGRILLHNGAYSITLQGLGISHLGLRSLRELGSGL 471
Db 407 ITGYLYISAMPDSLURDLSVFNLRIRGRILLHNGAYSITLQGLGISHLGLRSLRELGSGL 466

Qy 472 ALIHNTLHLCFVHTVPMDQLFRNPQHALLHTANRPEDE-CVGEGLACHQLCARHCWGPG 530
Db 467 ALIHNAHLCFVHTVPMDQLFRNPQHALLHSGNRPEDLCLVSSGLVCNSLCAHCHWGPG 526

Qy 531 PTQCVNCSQFLRGQECVEECRVLQGLPREYNARHCLPCHPCECQPNQSVTCFQPEADQC 590
Db 527 PTQCVNCSHFLRGQECVEECRVMKGLPREYVSDKRCCLPCHPCECQPNSSCTCFGEADQC 586

Qy 591 VACHAHYKDPFCVACRSGVKPDLSPYMPIKFPDEEGACQPCINCHTSCVDLDDKCPA 650
Db 587 AACHAHYKDSKSSVACRSGVKPDLSPYMPIKFPDEEGICQPCINCHTSCVDLDERGCPA 646

Qy 651 EQRASPLTSIVSAVVGILLVVLGVWFGLIKRQKIRKYTMRLQLQETELVELPLTPSG 710
Db 647 EQRASPTFIATVVGILLVVLGVWFGLIKRQKIRKYTMRLQLQETELVELPLTPSG 706

Qy 711 AMPNQAQMRILKETELARKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSIPKAN 770
```

```
Db 707 AMPNQAQMRILKETELARKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSIPKAN 766
Qy 771 KEILDEAYVMAGVSGPYVSRLLGLCLTSTVOLVTLQMPYGCCLLDHVRNRRGLSGQDLN 830
Db 767 KEILDEAYVMAGVSGPYVSRLLGLCLTSTVOLVTLQMPYGCCLLDHVRHRRGLSGQDLN 826
Qy 831 WCMQIAKMSYLEDRVLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKV 890
Db 827 WCVQIAKMSYLEDRVLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKV 886
Qy 891 PIKMALESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQ 950
Db 887 PIKMALESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQ 946
Qy 951 PICTIDVYMIWVKCWMIDSECRPRFRELVSFESFMRDPPQRFVVIQNEIDLGPASPLDSTF 1010
Db 947 PICTIDVYMIWVKCWMIDSECRPRFRELVSFESFMRDPPQRFVVIQNEIDLGPSSPMDSTF 1006
Qy 1011 YRSLLDEDDMGDLVDAEYLVPOQFFCPDAPCAGGMVHRRHSSSTRSCGGDLTLGLE 1070
Db 1007 YRSLLDEDDMGDLVDAEYLVPOQFFCPDPTGTSTAHRRHSSSTRSCGGELTLGLE 1066
Qy 1071 PSBEAPRSLAPSEAGSDVFDGLQMGAAKGLQSLPTHDPSPQLRYSDEPTVPLPSET 1130
Db 1067 PSBEAPRSLAPSEAGSDVFDGLAMGVTKGLQSLSPHDLSPQLRYSDEPTVPLPSET 1126
Qy 1131 DGYVAPLTCSPQPEYVNVQDVRPOPSPRGPPLPAAPAGATLERAKTSLSPGKGVWVDV 1190
Db 1127 DGYVAPLACSPQPEYVNVQSEVQPPPLTPGGLPPVRPAGATLERPKTSLSPGKGVWVDV 1186
Qy 1191 FAFGAVENPEYLTPOCGAAPPHPPPAFPAFONLYWDDPPERGAPPSTFKCTPTAE 1250
Db 1187 FAFGAVENPEYLVPRGTASPPHSPAFPAFONLYWDDQNSSEGGPPPNFEGTPTAE 1246
Qy 1251 NPEYLGLDVVP 1261
Db 1247 NPEYLGLDVVP 1257

RESULT 3
ERB2 MESAU
ID ERB2 MESAU STANDARD; PRS; 1254 AA.
AC Q6053;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RA MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
```

```

CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D16295; BAA03801.1; -.
CC HSSP: P11362; 1FGK
CC InterPro: IPR000494; EGFR_L_domain.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_pkinase.
CC InterPro: IPR004019; YLP_motif.
CC Pfam: PF00069; pkinase; 1.
CC Pfam: PF00757; Furin-like; 1.
CC Pfam: PF01030; Recep_L_domain; 2.
CC Pfam: PF02757; YLP; 2.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART: SM00261; FU; 3.
CC SMART: SM00219; TyrKc; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Proto-oncogene; Disease mutation.
CC SIGNAL 1 21 POTENTIAL.
CC CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
CC DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 653 675 POTENTIAL.
CC DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 158 368 CYS-RICH.
CC DOMAIN 472 644 CYS-RICH.
CC DOMAIN 720 987 PROTEIN KINASE.
CC NP_BIND 726 734 ATP (BY SIMILARITY).
CC BINDING 753 753 ATP (BY SIMILARITY).
CC ACT_SITE 845 845 BY SIMILARITY.
CC DISULFID 195 204 BY SIMILARITY.
CC DISULFID 199 212 BY SIMILARITY.
CC DISULFID 236 244 BY SIMILARITY.
CC DISULFID 240 252 BY SIMILARITY.
CC DISULFID 255 264 BY SIMILARITY.
CC DISULFID 268 295 BY SIMILARITY.
CC DISULFID 299 311 BY SIMILARITY.
CC DISULFID 315 331 BY SIMILARITY.
CC DISULFID 334 338 BY SIMILARITY.
CC DISULFID 511 520 BY SIMILARITY.
CC DISULFID 515 528 BY SIMILARITY.
CC DISULFID 531 540 BY SIMILARITY.
CC DISULFID 544 560 BY SIMILARITY.
CC DISULFID 563 576 BY SIMILARITY.
CC DISULFID 567 584 BY SIMILARITY.
CC DISULFID 587 596 BY SIMILARITY.
CC DISULFID 600 623 BY SIMILARITY.
CC DISULFID 626 634 BY SIMILARITY.
CC DISULFID 630 642 BY SIMILARITY.
CC MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).
CC VARIANT 658 658 V -> E (IN ONCOGENIC NEU).
CC VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
CC SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

```

```

Query Match 86.0%; Score 5870.5; DB 1; Length 1254;
Best Local Similarity 84.9%; Pred. No. 4.5e-311;
Matches 1088; Conservative 57; Mismatches 89; Indels 47; Gaps 4;

Qy 1 MELAAALCRWGLLLALLPPGAASSTOVCTGTDMLKRLPASPTHLDMLSHLYOGCVOVGNL 60
Db 1 MELAAACGGLLLALLSPGASGTQVCTGTDMLKRLPASPTHLDIVRHLYOGCVOVGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLAHNOVQVPLQRLRIVRGTQQLFEDNYALAVLDNG 120
Db 61 ELTYLPANATLSFLQDIQEVQGYMLAHSQVRHVPLOQLRIVRGTQQLFEDKYALAVLDNR 120
Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHNQOLA 180
Db 121 DPLDNVTATGRTPEGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHNQOLA 180
Qy 181 LTLIDTNRSPACHPCSPMKSGSCWGESSEDCQSLRTV-----C 220
Db 181 PVDIDTNRSPACHPCSPMKSGSCWGESSEDCQSLRTV-----C 220
Qy 221 AGGCARCKGPLPFNNFTVSWFLRVKVSASHLEKSHSDCLACLFHNSGICELHCPALVTY 280
Db 221 AGGCARCKGPLPFNNFTVSWFLRVKVSASHLEKSHSDCLACLFHNSGICELHCPALVTY 280
Qy 241 AGCT---GP-----KHSDCCLACLFHNSGICELHCPALVTY 274
Db 241 AGCT---GP-----KHSDCCLACLFHNSGICELHCPALVTY 274
Qy 281 NTDTFESMPNPEGRYTFGASCVTACPNYLSLTDVSGCTLVCPHLNQEVTAEDGTQRCCK 340
Db 275 NTDTFESMPNPEGRYTFGASCVTTCPNYLSLTDVSGCTLVCPHLNQEVTAEDGTQRCCK 334
Qy 341 SKPCARVCYGLGMEHLREVRANTSAIOBFGCKKIFGSLAFIPESDGPASNTAPLOP 400
Db 335 SKSCARVCYGLGMEHLRGARAITSAIQBAGCKKIFGSLAFIPESDGPSSGIAPLTP 394
Qy 401 EQLQVETLEETGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLSLQGLIGISWL 460
Db 395 EQLQVETLEETGYLYISAWPDSLHDLVSFQNLQVIRGRILHNGAYSLSLQGLIGISWL 454
Qy 461 LRSRLRELGSGLALIHNTLHLCFVHTVPDQLFRNPHQALLHTANRPEDECVGEGLACHOL 520
Db 455 LRSRLRELGSGLVLIHRNTHLCFVHTVPDQLFRNPHQALLHSGNPSEECGLKDFACVPL 514
Qy 521 CARGHCWGPQTQVNCQSFRLQCEVCECRVLQGLPREYVNAHCLPCHPECOQNGSV 580
Db 515 CARGHCWGPQTQVNCQSFRLQCEVCECRVLQGLPREYVNAHCLPCHPECOQNGSV 574
Qy 581 TCFGPADOCVACAHYKDPFPCVARGCPGVKPDLSYMPIWKFDEEGACACPCINCTHSC 640
Db 575 TCTGSEADOCVACAHYKDPFPCVARGCPGVKPDLSYMPIWKFDEEGACACPCINCTHSC 634
Qy 641 VLDDKGCPCAEQASPLTSIVSAVVGILLVVLVGVVFGILLIKRRQOKIRKYTMRRLLQET 700
Db 635 VLDDKGCPCAEQASPLTSIVSAVVGILLVVLVGVVFGILLIKRRQOKIRKYTMRRLLQET 694
Qy 701 ELVEPLTPSGAMPNQAMRILKETELRKVKVLSGSGAFGTVYKGIWIIPDGENVKIPVAIKV 760
Db 695 ELVEPLTPSGAMPNQAMRILKETELRKVKVLSGSGAFGTVYKGIWIIPDGENVKIPVAIKV 754
Qy 761 LRENTSPKANKETLDEAYVMAGVSPVSRLLGICLTSTVQLVTQMLPYPCLLDHVRNR 820
Db 755 LRENTSPKANKETLDEAYVMAGVSPVSRLLGICLTSTVQLVTQMLPYPCLLDHVRNR 814
Qy 821 GRIGSQDLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDE 880
Db 815 GRIGSQDLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDE 874
Qy 881 TEYHADGGKVPKWALESILRRRTHQSDVWSYGYTVWELMTFGAKPYDGIIPAREIPDL 940
Db 875 TEYHADGGKVPKWALESILRRRTHQSDVWSYGYTVWELMTFGAKPYDGIIPAREIPDL 934
Qy 941 LEKGERLPOPTCTTDVYIMVCKMIDSECPFRFRELVSFSESRMARDPQRFVVIQNEDL 1000
Db 935 LEKGERLPOPTCTTDVYIMVCKMIDSECPFRFRELVSFSESRMARDPQRFVVIQNEDL 994

```



```
Qy 957 VYIMVVKWMIDSECRPRFRELVSFGRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSL 1015
Db 943 VYIMVVKWMIDADSRKRELIIEFSKMDRDPQRYLVICQDERMHLPSPTDSNFYRALM 1002
Qy 1016 EDDDMGLVDAAEVLVPOQFFCDPPAGGAGMVHRRHSSTRSGGDLTLGLEPSEE 1075
Db 1003 DEEDMDVDVADEYLIPOQFFF-----SSPS 1028
Qy 1076 APRSPLAPSGAGSDVDGDLGMAAGKLOSLPHDPSLQRYSEDPTVLPSET--DGY 1133
Db 1029 TSRTPLLSLSATSN--NSTVACIDRNLGSCPIKEDSFQRYSSDTGTALTEDSIDTDF 1086
Qy 1134 VAPLTCSPQEVYNQDVRPQPPSREGPLPAARAGATLAKTLSPGKNGVVKQVFAF 1193
Db 1087 L-----PVPEYINQ-SVPRKRPAGSVQNVYHNPQPLN-----APSRDPHYQD--PH 1129
Qy 1194 GGAIVENPEYL-TPOGGAAPQHPPPAPAFADNLYYWDQ-----DP-----PER 1236
Db 1130 STAVGNPEYLTVQ-----PTCVNSTFDSAPHAQKQSHQISLDNPDYQODFFPKE 1180
Qy 1237 GAPPSTFKGTPTAENPEYL 1255
Db 1181 AKPNGIFKGS-TAENAYEL 1198
RESULT 5
EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RL (bek/KGFR) gene.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RL in mouse blastocysts during delayed implantation.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RN [5]
RP Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91212866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
```

```
RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eslinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X78987; CAA55587.1; -
DR EMBL; U03425; AAA17899.1; -
DR EMBL; X59698; CAA42219.1; -
DR EMBL; L06864; AAA53029.1; -
DR EMBL; Z12608; CAA78249.1; -
DR HSP; P11362; IFGK
DR MGD; MG1:95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
DR Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
DR SIGNAL 1 24
DR CHAIN 25 1210
DR DOMAIN 25 647
DR TRANSMEM 648 670
DR DOMAIN 671 1210
DR REPEAT 75 300
DR REPEAT 390 600
DR DOMAIN 1028 1071
DR DOMAIN 714 981
DR NP_BIND 720 728
DR BINDING 747 747
DR ACT_SITE 839 839
DR DISULFID 190 199
DR DISULFID 194 207
DR DISULFID 215 223
DR DISULFID 219 231
DR DISULFID 232 240
DR DISULFID 236 248
DR DISULFID 251 260
DR DISULFID 264 291
DR DISULFID 295 307
DR DISULFID 311 326
DR DISULFID 329 333
```



```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DFD2D2F5 CRC64;

```

Query Match 44.4%; Score 3031; DB 1; Length 1210;

Best Local Similarity 48.3%; Pred. No. 4.4e-157;

Matches 621; Conservative 171; Mismatches 362; Indels 132; Gaps 25;

```

Qy 11 LLLALLPPGAA--STQVCTGDMKLRLPASPETHLMRLHYQCQVQGNLEITYLPTN 68
Db 14 LTTALCAAGALEBKVCQGSNSRLTQGTGEDHPLSLQRMNVCVVLGNLEITYVORN 73
Qy 69 ASLSFLQDIQEVQVLIHNOVRQVLRIRVRGTQGLFEDNYALAVLDNGDPLNNTP 128
Db 74 YDLSFLQTIQEVAGVLIALTVERIPLNLIQIRGNALYENTYALISN----- 124
Qy 129 VTGASPGRLQLRLSLTEILKGVLTQRPOLCYQDTILWKDI-----FHKNOALATLI 184
Db 125 -YGNRTGLRELPRNLQELILIGAVRFSNNILCNMDTIQWRDIVQNVFMSNMDL--- 180
Qy 185 DTRNSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAAGCA-RCKGPLPFNNFTVSWFLR 243
Db 181 -QSHPSGPCDPSCPNGSCWGGGEENCQKTKIICAQCSHRCGRSPSD----- 230
Qy 244 VPKVSASHLE-----KHSDCIACLFHNSHGICELHCPALVTYNTDTFESMPNPGRY 295
Db 231 -----CCHNOCAAGCTGPRESDCLVCQKQFDEATCKDTCPPLMLYNTTYQMDVNPBGKY 285
Qy 296 TFGASCVTACPYNYLSTDVGSCTVCLPHNOEVTAEDGTQRCCEKSKPCARVCYGLGWEH 355
Db 286 SFGATCVKCPNRYVVTVDHGSVCVRACGPDYVEV-EEDGIRKCKKCDGPCRCVNCIGIGE 344
Qy 356 LREVRVTSANIQSFACKIIFGSLAFPESEFGDPSANTAPLOEQLOVPETLEEITGY 415
Db 345 FKDTLSINATNIRHKYCTALISGDLHLPLVAFKGDSPTRPPDLPREILTKTYKEITGF 404
Qy 416 LYISAWPDSLDLSVFQNLQVIRGILHNGAYSILTLQGLISWLGRLSLRGLSGLALIH 475
Db 405 LLIQAWPDNDTDLHAFENLEIRGTRKHQGFSLAVVGLNITSGLSLKEISDGDVILS 464
Qy 476 HNTLFCVHTVPDQLFRNPHQALLHTANPEDECVEGEGACHOLCARGHCWGPGTQCV 535
Db 465 GNRNLVCYANTINWKLFGTNPQTKIMNNAEKCKAVNHVCNPLCSSEGCGWGPEDCV 524
Qy 536 NCSQFLRQGCVEECRVLOGLPREYVYVNRHCLPCHPECPQNGSVTCFGPEADQCACAH 595

```

```

Db 525 SCQVSRGRCVEKCNILLEGEPREVFENSECQCHPECLPQAMNITCTGRGPDNCIQCAH 584
Qy 596 YKDPFFCVARPCSGVKPDLSTYMPWKFPDEEGACQPCPINCTHSCVLDLDDKGCAPQARAS 655
Db 585 YIDGPHCVKTCPCAGIMGENNTL-VWKYADANNVCHLCHANCTYGCAGCLOGCEVWPSPG 643
Qy 656 PLTSIVSAVGIILLVVLGVVFGI-LIKRROOKIRKYTRMRLLQETLVELPELTPSGAMPN 714
Db 644 KIPSATGIVGGLLFIVV-VALGIGLFMRRRHVRKRTLRLQLQRELVEPLTPSGEAPN 702
Qy 715 QAQWRILKETELRKVKVLGSGAFGVVYKGIIPDGENVKIPIVAKIVLENTSPKANKEIL 774
Db 703 QAHLRIILKETEFKKIKVLGSGAFGVVYKGLWIPGEKVKIPIVAKIELREATSPKANKEIL 762
Qy 775 DEAYVMAGVSPYVSRLLIGICTSTVQLVTQLMPPYGCLLDHHVRENRGLSGODLWNWCQ 834
Db 763 DEAYVMASVDNPHVCRLLIGICTSTVQLITQLMPPYGCLLDYVREHKDNIGSQYLLNWCQV 822
Qy 835 IAKOMSYLEDVRLVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETIYHADGGKVPKWK 894
Db 823 IAKGMNYLEDRRLVHRDLAARNVLKTPQHVKITDFGLAKLGAEEKEYHAEGGVKVPKWK 882
Qy 895 MALESILRRRTHQSDVMSYGVTYWELMTFGAKPYDGIPIAREIPDLLEKGERLPOPPICT 954
Db 883 MALESILHRIYTHQSDVMSYGVTYWELMTFGSKPYDGIPIASDISSILEKGERLPOPPICT 942
Qy 955 IDVTMIMVKMIDSECRPRELVSFBSRWARDPQRPVVIQ-NEDLGASPLDSTFYRS 1013
Db 943 IDVTMIMVKMIDADSRPRELILFBSRWARDPQRYLVIQGDRMHLPSPTDSNFYRA 1002
Qy 1014 LLEDDDDGLVDAAEYLVPOQFFCPDPAPGAGGMVHRHRSSTSSRGGGLTLGLERPE 1073
Db 1003 LMDEEDMEDVDAEYLIPOQGF-----NSPST----- 1031
Qy 1074 EEARPSPLASEGAGSDVFDGLGNGAAKGLQSLPTHDPSPLORYSEDTVPPLPSET--D 1131
Db 1032 ---SRTPLLSLSATS-----NSTVACINRNGSCRVEDAFQRYSSDPTGAVTEDNIDD 1084
Qy 1132 GYVAPLTCSPQEVYVNPQVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVF 1191
Db 1085 AFL-----PVPEVNO-SVPKRPAGSVONPVYINQPLHP-----APGRDLHYQN-- 1127
Qy 1192 AFGGAVENPEYL-TPQGGAAAPQHPFPAPSPAFONLYYWDQ-----DP-----P 1234
Db 1128 PHSNAVGNPPEYLTAQ-----PTCLSSGFSNLPALWIKGSHQMSLDNPDYQDDFFP 1178
Qy 1235 ERGAPPSTFKTPTAENPEYILGLDVP 1260
Db 1179 KETKPNGIFKG-PTAENAEYLRVAPP 1203

```

RESULT 6

```

ID_ERB4_HUMAN STANDARD; PRT; 1308 AA.
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
GN (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
OS ERBB4 OR HER4.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORM JW-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Roy L., Neubauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the

```

epidermal growth factor receptor family.";
Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
(2)
SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
TISSUE=Fetal brain;
MEDLINE=97476287; PubMed=9334263;
Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
Klagsbrun M.;
"A novel juxtamembrane domain isoform of HER4/ErbB4. Isoform-specific
tissue distribution and differential processing in response to
phorbol ester";
J. Biol. Chem. 272:26761-26768(1997).
CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC
CC -!- NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
CC
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: JM-A (SHOWN HERE) AND JM-B;
CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN HEART,
CC KIDNEY. IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
CC LUNG, SALIVARY GLAND, AND PANCREAS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L07868; AAB59446.1; -;
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3432; ERBB4.
DR MIM; 600543;
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00669; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD0000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 25
FT CHAIN 26 1308
FT DOMAIN 26 651
FT TRANSMEM 652 675
FT CYTOPLASMIC 676 1308
FT DOMAIN 186 334
FT DOMAIN 496 633
FT DOMAIN 718 985
PROTEIN KINASE.

| | | | |
|-----------------------|-------------------|---------------------|---|
| FT NP BIND | 724 | 732 | ATP (BY SIMILARITY). |
| FT BINDING | 751 | 751 | ATP (BY SIMILARITY). |
| FT ACT_SITE | 843 | 843 | BY SIMILARITY. |
| FT DISULFID | 189 | 197 | BY SIMILARITY. |
| FT DISULFID | 213 | 205 | BY SIMILARITY. |
| FT DISULFID | 213 | 221 | BY SIMILARITY. |
| FT DISULFID | 217 | 229 | BY SIMILARITY. |
| FT DISULFID | 230 | 238 | BY SIMILARITY. |
| FT DISULFID | 234 | 246 | BY SIMILARITY. |
| FT DISULFID | 249 | 258 | BY SIMILARITY. |
| FT DISULFID | 262 | 289 | BY SIMILARITY. |
| FT DISULFID | 293 | 304 | BY SIMILARITY. |
| FT DISULFID | 308 | 323 | BY SIMILARITY. |
| FT DISULFID | 326 | 330 | BY SIMILARITY. |
| FT DISULFID | 503 | 512 | BY SIMILARITY. |
| FT DISULFID | 507 | 520 | BY SIMILARITY. |
| FT DISULFID | 523 | 532 | BY SIMILARITY. |
| FT DISULFID | 536 | 552 | BY SIMILARITY. |
| FT DISULFID | 555 | 569 | BY SIMILARITY. |
| FT DISULFID | 559 | 577 | BY SIMILARITY. |
| FT DISULFID | 580 | 589 | BY SIMILARITY. |
| FT DISULFID | 593 | 614 | BY SIMILARITY. |
| FT DISULFID | 617 | 625 | BY SIMILARITY. |
| FT DISULFID | 621 | 633 | BY SIMILARITY. |
| FT MOD_RES | 1162 | 1162 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). |
| FT MOD_RES | 1188 | 1188 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). |
| FT MOD_RES | 1258 | 1258 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). |
| FT MOD_RES | 1284 | 1284 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). |
| FT CARBOHYD | 138 | 138 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT CARBOHYD | 174 | 174 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT CARBOHYD | 181 | 181 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT CARBOHYD | 233 | 233 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT CARBOHYD | 358 | 358 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT CARBOHYD | 410 | 410 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT CARBOHYD | 473 | 473 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT CARBOHYD | 495 | 495 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT CARBOHYD | 548 | 548 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT CARBOHYD | 576 | 576 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT CARBOHYD | 620 | 620 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT VARSPLIC | 626 | 648 | NGPTSHDCIYPTWGTGHTSLPQHA -> IGSSIEDICGLMD (IN ISOFORM JM-B). |
| FT SEQUENCE | 1308 AA; | 146807 MW; | 5E4AE80985D88761 CRC64; |
| Query Match | 42.6%; | Score 2905.5; | DB 1; Length 1308; |
| Best Local Similarity | 44.7%; | Pred. No. 3.1e-150; | |
| Matches 608; | Conservative 181; | Mismatches 376; | Indels 195; Gaps 31; |

| | | | | |
|----|-----|-----------------------|--|-----|
| QY | 9 | WGLLLALLPPGAA---- | STQVCTGTDMLKRLPASPTHLDMLRHLVQCGVQVQGNLELT | 64 |
| Db | 8 | WWWVSLVAAAGTVQPSDSQSV | CAGTENKLSUSDLEQQYRALRKYYENCEVWNGNLEITS | 67 |
| QY | 65 | LPTNASLSFLQDIOEVQGVLI | AHQVQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLN | 124 |
| Db | 68 | IEHNRDLSFLRSVREVTGVL | VALNQFVLPLENLRIIRGTGKLYEDRYALAIPLNVRKDG | 127 |
| QY | 125 | NTPYTGASPGGLRELQRLS | TEILKGGVLIQNPOLCYQDTILMKDIAPHKNQALTLI | 184 |
| Db | 128 | NP----- | GLOELGLKNLTEILNGGVYVDQNKFLCYADTIHWDIVRNWPNSNLTIV | 178 |
| QY | 185 | DTNRSRACHPCSPCKGSR | CWGESSEDCQSLTRTYCAGGC-ARCKGPLFNNFTVSWLR | 243 |
| Db | 179 | STNGSGGCRCHKSCGTG-RC | WGPTENHCOTLTRTYCAEOCDGRCYGPY-----V5----- | 227 |
| QY | 244 | VPKVSASHLE----- | KHSDCLACLFHNSGICELHCPALVITYNTDTFFSMPNPEGRY | 295 |
| Db | 228 | ---DCCHRECAAGCGSGPK | DTDCFACMFNDGACVTCQPTQFVFNPTFFOLEHFNAYK | 283 |
| QY | 296 | TFGASCVTACPNYLSLVG | STVGVCTLVCPHLNHOEVTAEDGTQRCSEKSKPCARVCYGLGMH | 355 |
| Db | 284 | TYGAFCVKCKPHNFV-VD | SSSCVRCAPSSKMEV-EENGKMKCKPCTDICKPCADGIGTGS | 341 |
| QY | 356 | LREVRVNTSANTQEPAGCK | KIFGSLAFLPESFDGDPASNTAPLOEQLOVFTLEITY | 415 |

Db 342 LMSAQTVDSNDIKFNCTKINGNLIFLVTGIHGDYNAIEADPEKLNVRFTVREITGF 401
Qy 416 LYISAMPDLSPLDSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLASRLASGLALIH 475
Db 402 LNIQSPPNMTDFSVSNLVTIGRVLYSGLSLILKQOGITSLQFQSLAKEISAGNIYIT 461
Qy 476 HNTLFCFVHTVPDQQLFRNPHQALLHTANRPEDECVEGGLACHOLCARGHCWGPQPQCV 535
Db 462 DNSNLCCYHTTNTTIFSTINQIRIVRDNRAENCTAEGMVCNHLSSDGCWGPDPQCL 521
Qy 536 NCSQFLRGQBCVEECVLOGLPREYVNAHCLPCHPECOPI-ONGSVTCFCPEADQCVACA 594
Db 522 SCRRFSGRICISCNLYDGEFREFNGSICVCDPOCEKVEDGLLTCGPGDPCNTKCS 581
Qy 595 HYKDPFCFCVARGSGVKPDLSPIMPKFPDEEGACOPCPINCHTSHCSVDLDDKGC----- 648
Db 582 HFKDGFCNVEKCPDGLQANSF--IPKADPDRECHPCNCTGCGNPTSHDCIYYPT 639
Qy 649 -----PAEQRASPLTISVNAV-GILLVVVLGVVFGILIKRQOKIRKYMRRLLQTEL 702
Db 640 GHSTLPQHAR-TPL--IAAGVIGLFLIVLGLTFVAVYVRKSGIK-KKRALRRFL-ETEL 694
Qy 703 VEPLTSGAMPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAKVLR 762
Db 695 VEPLTSGTAPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGETVKIPVAKILN 754
Qy 763 ENTSPKANKEILDEAYVMAGVGPVYSRLGICLTSTVQLVTQMLPYGCLLDHVRNRR 822
Db 755 ETTGPKANVEFMDEALIMASMDPHLVRLVGLVCLSPITQVLQTMHPHGCLLEYVHEKON 814
Qy 823 LGSQDLNLCMQIAGKMSYLEDVRLVHRDLAARNLVKSPNNVKITDFGLARLLDIDETE 882
Db 815 IGSQDLNLCMQIAGKMSYLEDVRLVHRDLAARNLVKSPNNVKITDFGLARLLDIDETE 874
Qy 883 YHADGKVPKMALESILRRRTHOSDVMSYGVYVWELMTFCGAKPDGIPAREIDPLE 942
Db 875 YHADGKVPKMALESILRRRTHOSDVMSYGVYVWELMTFCGAKPDGIPAREIDPLE 934
Qy 943 KGERLPQPPCTIDVTVMYKMWIDSECRPRELVSFSEMRAPQREWVQIONED-LG 1001
Db 935 KGERLPQPPCTIDVTVMYKMWIDSECRPRELVSFSEMRAPQREWVQIONED-LG 994
Qy 1002 PASPLDSTYRSLLEDGDLVDAREYLVQOQFCPDPAAGAGMHHRRSSSTRSG 1061
Db 995 LPSNDSKFFQNLLEDLEDMDABEYLVQ-PAFNIPPP-----IYTSRARDISNRS- 1046
Qy 1062 GGDLTGLSPSEEAERS-----PLAP-SEGAGSDVF 1092
Db 1047 -----EIGHSPPPAYTWSGNQFVYRDGGFAAEGSVYPRAPTSTIPEAPVAGATAEIF 1102
Qy 1093 DGDGLGAAKGLSLPHTDPSPLQRYSEDPVPLPS-----ETDGYVAPLTCSPQPEY 1145
Db 1103 DDCSCNGTLRKVPVPHVQEDSSQRYSDPTVPAPERSPRGELDEGYMTMDKPKQEY 1162
Qy 1146 VNQPDVVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKGVKVDVFAFGAVENPEYLT 1205
Db 1163 LNPVE-----ENPFVSRR-----KNGDLQ-----ALDNPEYHNA 1191
Qy 1206 QGGAAPQPHPPA-----FSPAENLYYWDQ 1231
Db 1192 SNG-----PPKAEDYVNEPLYNTFANTLGAKEYLKNILSMPEKAKAFDNPYWNH 1245
Qy 1232 DPPERGA--PPSTFKGTPT-----AENPEYL 1255
Db 1246 SLPPRSTLQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 7
ERB4 RAT
ID ERB4 RAT
AC Q62956; Q922N7; PRT; 1308 AA.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Balliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041838; AAC08899.1; -
CC EMBL; U52531; AAC53051.1; -
CC HSP; P11362; LFGLK.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00059; pkinase; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC Pfam; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25
FT CHAIN 26 1308
FT DOMAIN 26 651
FT TRANSMEM 652 675
FT DOMAIN 676 1308
FT DOMAIN 186 334
FT DOMAIN 496 633
FT DOMAIN 718 985
FT NP_BIND 724 732
FT BINDING 751 751
FT ACT_SITE 843 843
FT DISULFID 189 197
FT DISULFID 213 205
FT DISULFID 213 205
FT DISULFID 217 229
FT DISULFID 230 238
FT DISULFID 234 246
FT DISULFID 249 258
FT DISULFID 262 289
FT DISULFID 293 304
FT DISULFID 308 323
FT DISULFID 326 330
FT DISULFID 503 512
FT DISULFID 507 520
FT DISULFID 523 532
FT DISULFID 536 552
FT DISULFID 555 569
FT DISULFID 559 577
FT DISULFID 580 589
FT DISULFID 593 614
FT DISULFID 617 625
FT DISULFID 621 633
FT MOD_RES 1162 1162
FT MOD_RES 1188 1188
FT MOD_RES 1258 1258
FT MOD_RES 1284 1284
FT CARBOHYD 138 138
FT CARBOHYD 174 174
FT CARBOHYD 253 253
FT CARBOHYD 358 358
FT CARBOHYD 410 410
FT CARBOHYD 473 473
FT CARBOHYD 495 495
FT CARBOHYD 548 548
FT CARBOHYD 576 576
FT CARBOHYD 620 620
FT CONFLICT 1062 1062
FT CONFLICT 1080 1082
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match
Best Local Similarity 44.5%; Score 2886; DB 1; Length 1308;
Matches 606; Conservative 188; Mismatches 385; Indels 182; Gaps 31;

Qy 1 MELA-ALCRWGLLL--ALLPPGAASCTCTGDKMLRLPASPTHDLMLRLHYGQCQVWQ 57
Db 1 MKLATGLWWSLLVAARTVQPSASQSVCACTENKLSLSLSLEQYRALRYKYEYCEVVM 60

Qy 58 GNLELTLPNTWASLFLODIOEQVYLIANQVQPLQLRLIRVGTQLFEDNYALAVL 117
Db 61 GNLEITSIEHRDLSFLRSREVYGYVLVALNQRYLPLENLRIINGTKLYEDRYALAIF 120

Qy 118 DNGDPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQNPQLCYODTILWKDFIHKQN 177
Db 121 LNYRKGNF-----GLQELGLKMLTEILNGGVYVDQNKFLCYADTIHWQDILVRNPW 171

Qy 178 QALTLTIDTNRSRACHPCSPMKGSRGCMGESSEDCQSLTRTVCAAGC-ARCKGPLPNNF 236
Db 172 PSNMTLVSTIGSSCGGRCHKSKTG-RCWGPTEHNHQTILTRTVCAQCQGRGCVGY- 225

Qy 237 TVSFMLRPVKVSAHLE-----KHSDCLACLHFNHSGICELHCPALVYNTOTFSM 288
Db 226 -VS-----DCCHRECAGCGSKGDKTDCFAKMFNDGACVTCQPTQFVYNTPTFOLE 276

Qy 289 PNPEGRYTFGASCVTACPNYVLSLTDVGSCTLVCPHLNQEVTAEDTCRCKSKCARVC 348
Db 277 HNFNAKYTYGAFCKVKCPHNFV-VDSSCVTRACSSKMEV-EENIGKMKCPCTDLCPKAC 334

Qy 349 YGLGMEHLREVRVAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQVRET 408
Db 335 DGICTGSLMSAQTVDSNIDKFNCTKINGNLIFLVTGIGHGDPYNAIDAIDPEKLVNFR 394

Qy 409 LEEITGYLYISAWPDSLPLDSVFONLQVIRGRIHNGAYSUTLQGLGTSWGLRSLREL 468
Db 395 VREITGFLNIQTWPPNMTDFSVFNLVTIGRVLYSGLSLLILKQOGITSLQFSLKETS 454

Qy 469 SGLALIHNNTHLCFVHTVPMDLPFNPHQALLHTANRPEDECVGGLACHOLCARGHCWG 528
Db 455 AGNIYITDNSNLCTYHTINWTLFTSTVQRIIVIRDNRAENCTABGMVCHLUCSNDGCGW 514

Qy 529 PGPTQCVNCSQFLRGQECVEECRVLQGLPREYVVARHCLPCHPECP-QNGSVTFCGPEA 587
Db 515 PGDQCLSCRRFSRGKICIESCNLYDGEFRFENGSI CVCDSQCEKXEDGULLTCHGPGP 574

Qy 588 DQCVACAHYKDPFPCVARGSPGVKPDLSYMPIWKPDEEGACQCPINCTHSCVOLDKNG 647
Db 575 DNCTKCSHFQDGNPCVKECPDVLQGANF--IFKYADQDRECHPCHPNCQTCGNGPTSHD 632

Qy 648 C-----PAEQRASPLTSIVSAV-GILLVVVLGVVFGILIKRQOKIRKYVWR 695
Db 633 CIYYPWTGHSHTLPQAR--TPL--TAAGVIGELFLIVMALFAVYVRRRSK-KKRALRR 688

Qy 696 LLOETELVEPLTPSGAMPNQAQMRILKETELRKVKVLSGSGAGFYVYKGIWIPDGNVXIP 755
Db 689 FL-ETELVEPLTPSGTAPNQAQRLKETELRKVKVLSGSGAGFYVYKGIWPEGETVKIP 747

Qy 756 VAIKVLRENTSPKANKEILDEAYVMAGVSGSPYVSRLLGICITSTVQLVTQMLPYCGLDH 815
Db 748 VAIKILNETTGPKANVEFMDEALIMASVDHPLVRLVCLSPITQVLTQMLPHGCLLEY 807

Qy 816 VRENRLGSDLLNMCQIAKMSYLEDVLRVHRDLAARNVLKSPNHVKITDFGLARL 875
Db 808 VHEKDNIGSOLLNWCQIAKGMVLEERLVRDLAARNVLKSPNHVKITDFGLARL 867

Qy 876 LDIDETEHADGKVPVKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAR 935
Db 868 LEGDEKEYNADGGKMPVKWMALECIHYRKFTHQSDVMSYGVTVWELMTFGKPYDGI PTR 927

Qy 936 EIPDLLEKGERLPQPPICTIDVYIMVKCMWIDSECRPRFRELSEFSEFMRMARDPQFVVI 995
Db 928 EIPDLLEKGERLPQPPICTIDVYIMVKCMWIDADSRPKFELAAEFSEFMRMARDPQFVVI 987

Qy 996 QNED-LGPASPLDSTFYRSLLEDDMDGLVDABEYLVPOQGFCCDP----- 1041
Db 988 QGDDRMKLPSPNDSKFFQNLDEEDLEDDMDABEYLVLP-QAFNIPPPITYSTRTRIDSRS 1046

Qy 1042 -----APGAGGMVHRHRSSTRSGGDLTLGLPESEERAPRPLAPSEGAGSDV 1091
Db 1047 EIGHSPPPAYTPMSGSQFVYDGGFATQCG--NPMPTATSTIPEAPVA--QCATAM 1101

Qy 1092 FDGLDGMGAAGLQSLPETHDPSLQRYSEDTVPPLPS-----ETDGYVAPLTCSPQPE 1144
Db 1102 FDDSCCNGTLRKPVVPHVQEDSSQRYSDPTQRYAPERNPRAELDEEGYMTPMHDKPQOE 1161

Qy 1145 YVNPQDVRPQPPSPREGFLPAARAGATLERAKTSLSPGKNGVVDVFAFGGAVENPEYLT 1204
Db 1162 YLNPE-----ENPFVSR--KNGDLQ-----ALDNPEYHS 1190

Qy 1205 PQGGAAPQHPHPPA-----FSPAFDNLVYWD 1230


```
Db 410 VIMMPEMTSLSFQWLEIIRGRTTFSRGFSFVVQVVRHLQWLGRLSLKEVSAGNVILK 469
Qy 476 HNTLFCFVHTVPDOLERNPHQALLHTANRPEDECVEGLACHOLCARGHGWGPGTQCV 535
Db 470 NTQLRQVANTINRRLFRSDQSLEYDART-----ENOTCNNECEDGCGWGPPTMCV 522
Qy 536 NCSOFLRQGBCEBRCVRLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADOCVACAH 595
Db 523 SCLHVDGGRGCVASCNLQGEPREAOVDGRVCQCHQECQLVQTDLSLTCYCGPANCSSAH 582
Qy 596 YKPPFCVACRCPGVKPDLSYMPIWKFPEDEGACQPCPINCSTHSCVLDLDDKGCBAERAS 655
Db 583 FQDGPQICPRPHGILGDDTL-IMKYADKMGQCPQCHQCTQCGSGFSLGSGRGD-IVS 640
Qy 656 PLTSIVSAGVILLVAVGVVGLIKRRQOKIRKTYMRRLLQELTELVEPLTPSGAMPNQ 715
Db 641 HSSLAVGLVGLLITVIVALLVLLRRRIK-RKTIIRCLLQKEKELVEPLTPSGAPNQ 699
Qy 716 AQMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPPKANKELD 775
Db 700 AFLRIKETEFKDRVLGSGAFGTYYKGLWNPDCENIRIPVAIKVLRATSPKVNQEVLD 759
Qy 776 EAYMAGVGSFYVSRILGICLTSTVOLVTLQMPYGCCLLDHVNRGRGLSGODLLNWCQOI 835
Db 760 EAYMASVDHPVHVCRLGICLTSAVLVTLQMPYGCCLLDVYVQHOERICGQWMLNWCQOI 819
Qy 836 AKGMSYLEDVRLHVDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPKWM 895
Db 820 AKGMYLEERHLVHVDLAARNVLKSPNHVKITDFGLSKLLTADKEKYQADGGKVPKWM 879
Qy 896 ALESILRRRTHQSDVMSGYVYVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTI 955
Db 880 ALESILQWTVTHQSDVMSGYVYVWELMTFGSKPYDGIPIAREIASVLENGERLPPOPICTI 939
Qy 956 DVYIMVWKMWIDSECPREFELVSESRHARDPQRFVQNEIDLGPASPLDSTFYKSL 1015
Db 940 EYVIMLKWMWIDSPREFELVSESRHARDPQRFVQNEIDLGPASPLDSTFYKSL 996
Qy 1016 EDDMGDLVDAEYLVQOQFCFDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEE 1075
Db 997 SSD--DVDAVELLYPKRI-----NQGS----- 1020
Qy 1076 APRSLAPSEGAGSDVFDGLMGAKGLOSPLTHDPSPLQRYSEDTV-PLPSETGYV 1134
Db 1021 ---EPCIPPTGH-----PVRENSITLNRISDPTQNALEKDLGDH- 1056
Qy 1135 APLTCSQPEYVNPQDVRPQP-----PSPRE-----GPLP-AARFAGATLERAKTLSP 1191
Db 1057 -----EYVNPQGETSRSLSDIYNPNYEDLTGCGVPVSLSSQEAETNFSRPEYLNT 1107
Qy 1182 GKNGVWKDVFAFGAVENPEYLTQGGAAAPQHPPPAFSPAFDNLVYWDQDPPRGPAPS 1241
Db 1108 NQNSL---PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTG 1142
Qy 1242 TFKGTPTAENPEYLG 1256
Db 1143 NGMFLPAENLEYLG 1157
```

RESULT 9

```
ERB3_HUMAN
ID ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
GN (c-erbB3)
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=90083234; PubMed=2687875;
Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
"Isolation and characterization of ERBB3, a third member of the
ERBB/epidermal growth factor receptor family: evidence for
overexpression in a subset of human mammary tumors.";
Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
[2]
SEQUENCE FROM N.A.
MEDLINE=90311312; PubMed=2164210;
Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
Todor G.J., Shoyab M.;
"Molecular cloning and expression of an additional epidermal growth
factor receptor-related gene.";
Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
[3]
SEQUENCE FROM N.A. (SHORT FORM).
TISSUE=Placenta;
MEDLINE=93282822; PubMed=7685162;
Kato M., Yazaki Y., Sugimura T., Terada M.;
"c-erbB3 gene encodes secreted as well as transmembrane receptor
tyrosine kinase.";
Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
SECRETED (SHORT FORM).
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/ebis-sib.ch/announcement/
or send an email to license@sib-sib.ch).
-----
EMBL; M29366; AAA35790.1; -
EMBL; M34309; AAA35979.1; -
EMBL; S61953; AAB28935.1; -
PIR; A36223; A36223.
HSSP; P11362; 1FGK.
Genew; HGNC:3431; ERBB3.
MIM; 190151; -
InterPro; IPR000494; EGFR_L_domain.
InterPro; IPR000719; Euk_pkinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR01245; Tyr_pkinase.
Pfam; PF00069; pkinase; 1.
Pfam; PF00757; Furin-like; 1.
Pfam; PF01030; Recep_L_domain; 2.
ProDom; PD000001; Euk_pkinase; 1.
SMART; SM00261; FU; 3.
SMART; SM00219; TyrcK; 1.
PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
```


RESULT 10
ERB3 RAT STANDARD; PRT; 1339 AA.
AC Q62759; Q62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERBB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
recombinant protein.";
RL Gene 165:279-284(1995).
RN [2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuroligins and their putative receptors, ErbB2 and
ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; U29339; AAC28498.2; -;
DR EMBL; U25330; AAC53050.1; -;
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR PRINTS; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;

KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339
FT DOMAIN 20 643
FT TRANSMEM 644 662
FT DOMAIN 663 1339
FT DOMAIN 183 259
FT DOMAIN 707 964
FT NP_BIND 713 721
FT BINDING 740 740
FT ACT_SITE 832 832
FT DISULFID 186 194
FT DISULFID 190 202
FT DISULFID 210 218
FT DISULFID 214 226
FT DISULFID 227 235
FT DISULFID 231 243
FT DISULFID 246 255
FT DISULFID 259 286
FT DISULFID 290 301
FT DISULFID 305 320
FT DISULFID 323 327
FT DISULFID 500 509
FT DISULFID 504 517
FT DISULFID 520 529
FT DISULFID 533 549
FT DISULFID 556 573
FT DISULFID 576 585
FT DISULFID 589 610
FT DISULFID 613 621
FT DISULFID 617 629
FT CARBOHYD 126 126
FT CARBOHYD 250 250
FT CARBOHYD 353 353
FT CARBOHYD 408 408
FT CARBOHYD 414 414
FT CARBOHYD 437 437
FT CARBOHYD 469 469
FT CARBOHYD 522 522
FT CARBOHYD 566 566
FT CARBOHYD 616 616
FT CONFLICT 1028 1028
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BPDF1 CRC64;
Query Match 33.3%; Score 2271.5; DB 1; Length 1339;
Best Local Similarity 40.0%; Pred No. 7.7e-116;
Matches 519; Conservative 168; Mismatches 433; Indels 177; Gaps 36;
Qy 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMLRLPASPETHIDMLRHLHYQGCGQWQGN 59
Db 7 LQVLC---FLLSLARGSEMGNSQAVCPGTLNGLSVTGDADNQYQTLVKLYEKCEVVMGN 62
Qy 60 LEITYLPTNASLSFLQDIQIEVOGYVLIHNOVROVPLQRLIRIVRGTLQFDENYALAVLDN 119
Db 63 LEIVTGHNADLSFLQIWEIVRTGTVLVNMFNFEVPLNLRVVRGTQYDVGKFAIFVN-- 120
Qy 120 GDPINNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPCQYQDTILWKDIFHKNQL 179
Db 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNKLCMDTIDMDRIVRVR--- 170
Qy 180 ALTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQLSTRITVCAGGC-ARCKGLPLFPNFTV 238
Db 171 GAELVVKNGANGCPPCHEVCKG-RCWGPDPDCCQLTKTICAPQCGRCFCFNP----- 223
Qy 239 SFWLRLVPKVSASHLE-----KHSCLACLHPNHSGLCELHCPALVYNTDTFESMPN 290
Db 224 -----NQCHDECAGGCGSQDTCFACRRFNDSGACVPRCPPEPLVYNKLTFOLEPN 275
Qy 291 PEGRYTFGASCVTACPVNYLSTVGSCITLCPHLNQSVTAEDGTQRCCKSKPCARVCYG 350
Db 276 PHTKYQYGGVCVASCAPHNFV-VQOTFCVRACPPDKMEVD-KHGLKMCPECGGLCPKACEG 333
Qy 351 LGMHILREVRAVTSANIQEFAGCKIFGSLAFLPSPDGPASNTAPLQPEQLQVFTLE 410


```
Db 334 TCSSG--SRYTQVDSNIDGFVNCTKILGNLDELITGLNVDPMWHKIPALDPEKLNVTYR 391
Qy 411 EITGYLYISAWPDSLPDLVSFONQVIRGRILHNGAYS-LTQGLGIGWGLRSLRGLGS 469
Db 392 EITGYLNTOSWPHHNFVSFNLTITIGRSLYNRGFSLLIMKNLNTVSLGRFSLKEISA 451
Qy 470 GLIALIHNTHLCFVHTVPMDQLFRNPQALLHTA-NRPEDECYGEGLACHOLCARGHCWG 528
Db 452 GRVYISANQOLCVHHSLNWTRLLRGPSEERLDIKYDRPLGCELAEGKVCDDLPCSSGCGWG 511
Qy 529 PPTOCVNCQSLRGQCEVCEBVLQCLPREYVNAHCLPCHPECOPOQNGSVTCFPEAD 588
Db 512 PGPQGLSCRNYSREGVCTVCHNCFLQGEPRFVHEAQCFSCHPECLMPEGTSTCNGSGD 571
Qy 589 QCVACAHYKDPFCVACRCPGKVPDLSPYMPKPFDEEGACOPCPNCTHSC--VLDLDK 646
Db 572 ACARCAHFRDGHCVNSCHPGLG--AKGIYKYPDAQNECRPCHECTQCGNGPELQDC 629
Qy 647 GCPAQRASPLTSISAVVGVILLVVLGVVFGILIKRQOKIR-KYTMRLLLQETELVEP 705
Db 630 LGOAEVLMSKPHLVIATVVG--LAVILMILGGSFLYWRGRRIQNKRAMRYLERGESIEP 687
Qy 706 LTPSGAMPNQAOMRILKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRENT 765
Db 688 LDPS-EKANKVLARIFKETELRKVKVLGSGVFGVTHKGIWIPGESIKIPVCIKVIEDKS 746
Qy 766 SPKANKEIIDEAYVMAGVSPVYRLLGICLTSTVOLVTLQMPYGLLDHVRNRRGLGS 825
Db 747 GRQSQAVTDHMLAVGSLDHAHIVRLGLCPGSSLQVLVQYPLGSLLDHVRKHRRITLGP 806
Qy 826 QLLNWCMIQIAGMSVLEDRVLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETFYHA 885
Db 807 QLLNMGVQIAGMYLLESHSVHRLDALRNVMKSPSQVQVADFGVADLLPDDKQLLH 866
Qy 886 DGGKVPKMALESILARRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKE 945
Db 867 SEAKTPKMALESIHFGKYTHQSDVMSYGVTVWELMTFGAEYAGLRLAEPDLLEKE 926
Qy 946 RLPPQPICTIDVYMWKCMIDSECRPRFRELVSFSSRMARDPQRFVVIQNEIDLGPSP 1005
Db 927 RLQAQPICTIDVYMWKCMIDENIRPTFKELANEFTMRADPPRYLVIKRAS-GPGTP 985
Qy 1006 LDSTFYRSLLEDDMDGLDAEYLVPOQGFPCPDPAAGGGMVHRRHSSTRSGGDL 1065
Db 986 --PAAEPSVLTKEL-----QEALEPEL-----DL 1009
Qy 1066 TLGLEPSEE-----EAPRSLAPSEG-----AGSDVFGDGLG 1097
Db 1010 DLDLAEEBGLATSLGSLSLPTGTLTRPRGSQLSPSSGYPMNQSSLGEACLDL 1069
Qy 1098 MGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVV-----APL-----TC-----SP 1141
Db 1070 GCREQFSRPISLH-PIPRGR-----PASESEGHVTCSEALQEKVSVCSRSRSP 1121
Qy 1142 QPE-----YVQPDVVRPPSPREGP-----LPAARPATATLERAKTLSP-GKNG 1185
Db 1122 RPRGDSAVHSQRSHLLTPVTPLSPGLEEDGNGYVMPDTHLRGASSSREGTLSSVGLSS 1181
Qy 1186 VV-----KDVAFGGAIVENPEYLTPOGGAAPQPHPP 1216
Db 1182 VLGTEEDD-----EEYEMNRKRGSP-PRPP 1209
```

RESULT 11

```
EGFR_DROME
ID EGFR_DROME STANDARD: PRT; 1426 AA.
AC P04412; O61601; QW260; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (torpedo protein) (Drosophila relative of ERBB).
```

```
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL Genetids 137:531-550(1994).
RN [2]
RP REVISIONS.
RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chexry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ithegawam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.G., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
```

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*;"
RL Science 287:2185-2195(2000).
[7]
RN SEQUENCE OF 959-1078 FROM N.A.
RP STRAIN=Daekwanryeong;
RC MEDLINE=85137938; PubMed=2983232;
RX Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RA "A *Drosophila* genomic sequence with homology to human epidermal
RT growth factor receptor;"
RL Nature 314:178-180(1985).
[8]
RN SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS.
RX MEDLINE=92038942; PubMed=1936959;
RA Raza E., Schejter E.D., Shilo B.Z.;
RT "Interallelic complementation among *DER/fib* alleles: implications for
RL the mechanism of signal transduction by receptor-tyrosine kinases;"
RL Genetics 129:191-201(1991).
[9]
RN REVIEW
RX MEDLINE=97248481; PubMed=9094709;
RA Perrimon N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the *Drosophila*
RL EGF receptor;"
RL Cell 89:13-16(1997).
CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAP-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOTEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF052754; AAC08536.1; JOINED.
CC EMBL; AF052753; AAC08536.1; JOINED.
CC EMBL; AF052754; AAC08535.1; JOINED.
CC EMBL; AF052752; AAC08535.1; JOINED.
CC EMBL; K03054; AAA51462.1; JOINED.
CC EMBL; K03417; AAA51460.1; JOINED.

DR EMBL; K03416; AAA50965.1; JOINED.
DR EMBL; K03418; AAA51461.1; JOINED.
DR EMBL; AF109077; AAD26134.1; JOINED.
DR EMBL; AF109078; AAD26132.1; JOINED.
DR EMBL; AF109082; AAD26132.1; JOINED.
DR EMBL; AF109078; AAD26133.1; JOINED.
DR EMBL; AF109084; AAD26133.1; JOINED.
DR EMBL; AF109079; AAD26130.1; JOINED.
DR EMBL; AF109081; AAD26130.1; JOINED.
DR EMBL; AF109079; AAD26131.1; JOINED.
DR EMBL; AF109083; AAD26131.1; JOINED.
DR EMBL; AF109080; AAD26135.1; JOINED.
DR EMBL; AE003454; AAF46732.1; JOINED.
DR EMBL; K02293; CAA26157.1; JOINED.
DR EMBL; X78920; CAA55523.1; JOINED.
DR EMBL; X78918; CAA55521.1; JOINED.
DR EMBL; X78919; CAA55522.1; JOINED.
DR PIR; A00640; GQFFE.
DR HSP; P11362; IFGK.
DR FlyBase; FBgn0003731; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recept_L domain; 2.
DR PRINTS; PD00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW Developmental protein.
FT SIGNAL 1 30
FT CHAIN 31 1426
FT DOMAIN 31 868
FT TRANSMEM 869 889
FT DOMAIN 890 1426
FT DOMAIN 938 1198
FT NP_BIND 944 952
FT BINDING 971 971
FT ACT_SITE 1063 1063
FT MOD_RES 902 902
FT PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
Query Match 27.6%; Score 1883; DB 1; Length 1426;
Best Local Similarity 32.1%; Pred. No. 9.8e-95;
Matches 461; Conservative 184; Mismatches 433; Indels 360; Gaps 43;
Qy 24 QVCTGDMKURLPASPEHLDMLRHLYQGCVVQGNLELYLPT-NASLSFLQDIQEVQG 82
Db 100 KICIGTKSRSLSPSNKEHHYRNLDRYTNCTYDGNLKLTLWLPNENLDSFLDNIREVTG 159
Qy 83 YVLIANNQVQVPLQRLIRIVRGTLF-----EDNYALAVLDNGDPLNTPPTVTCASPGGL 137
Db 160 YILISHVDVKVFPKLIIRGRTLFLSLSVEEEKYALFV-----TYSRM 203
Qy 138 RELQLASRLTEILKGGVLIQRNPOLCYQDTTLTKWDIFHKNNQLALTLIDNRSRACHPCSP 197
Db 204 YTLLEIPDLRDVLNGQVGFHNNYLNCHWRTTIQNSEIYVNGTDAYNYDFTAPERCPCKE 263
Qy 198 MCKSGRCWGESSEDQSLTRTVCAAGCA--RCKGPLFPNNFTVSWFLRVPKVSASHL--- 252
Db 264 SCTHG-CWSEGPKNCKFKSLTCSPOCAGGRCYGPKP-----RECCHLPCA 308
Qy 253 -----EKHSCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGYTTCACVTCACP 307
Db 309 GGCTGPTQKDCIACKNFFDEAVSKBECPPMRKRYNPTTYVLETPNPEGYATGATCVKECP- 367
Qy 308 NYLSTDVGSCTLVCPLNHQBVTAEADGTQRCCKSKPCARVCYGLGMHLEHREAVTSANI 367

```
Db 368 GHLLRNGACVRSQPOKMDKGE-----CVPNGCPKTCPGVTVLH-----AGNI 414
Qy 368 QEFAGCKIFGSLAFESFDG--DPASNTA-----PLQEOLOVFETLEETGLYLTIS 419
Db 415 DSFRNCTVIDGNIRILDQTFSGFDQVYANYTWGPRVPLDPEREVEFSTVKEITGYLNTIE 474
Qy 420 ANPDSLPLDSVFONQVIRGRILHNGAY-SLTLOGLGISWGLRSLRSLGSLALHHNT 478
Db 475 GTHPOPRNLISYFRNLETIHGRQLMESFAALAIVKSSLSYLEMRNLKQISSGSSVVIQHRN 534
Qy 479 HLCFFVHTVPWDOLFERNPHQALLHTANRPEDECVGEGLACHQLCARGCHGPGPTQCVNCS 538
Db 535 DLQYVSNIRWPAIQEPEKQVWVWENLRADLCEKNGTICSDQNEGCGWAGTDQCLTCK 594
Qy 539 QFLRGOECVEECRVLOGLPREYVNAHCLPCHPEQOPQNGSVTCFGEADQCVACAHYKD 598
Db 595 NFNFGTICADCGYISNAYK--FDNRCKICHPECR-----TCNGAGADHCOECVHVED 646
Qy 599 PFFCVARCP-----SGVK-----611
Db 647 GQHCVSECPKNKYNDRGVCRECHATCGCTGPKDTIGIGACTTCNLAIINNDATVKRCLL 706
Qy 612 -----PDLSPMPWKF--PDEGACOP-----CPI-----NCTH- 638
Db 707 KDDKCPD-CY--FWEVHPQEOGSLKPLAGRAVCRKCHPLCELTNGYHGEQVCKTHY 763
Qy 639 -----ASPLTS-----SC-----VDLDDKG-----647
Db 764 KRREQETECPADHYTDEBQECFQRHPBCNGCTGPGADDCKSCRNFKLFDANETGPYVN 823
Qy 648 -----CPAEQR-----ASPLTS-----IVSAVVGILLVV 671
Db 824 STMFNCTSKCPLEMRHVNQYTAIGPYCAASPPRSKITANLDVNMIFITGAVLVPTIC 883
Qy 672 VLGVVFGILIKRQOKIRKYT--MRLLOTELVELTPSGAMPQAOAMRILKTELKV 729
Db 884 ILCVV--TYICQKQKAKKETVMTWALSGCDSQSPLEPSNTGALCKURIVKDAELRG 941
Qy 730 KYLGSGAGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVSPYVS 789
Db 942 GVLGMAFGFRVYKGVWVPEGENVKIPVAIKELKSTGAESSEFEAREAYIMASEEHVNL 1001
Qy 790 RLGLCLTSTVOLVTLQMPYGCLLDHVRNRENRGLSGQDLNMCQIAKMSYLEDRVLVH 849
Db 1002 KLLAVCNSSQMLITQLMPLGLCLDYVRNRDKIGSKALLNWSQTAKGMSYLEKRLVH 1061
Qy 850 RDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRBRTQS 909
Db 1062 RDLAARNVLVQTPSLVKITDFGLAKLLSDSNEYKAAGGMPIKWLALCEICRNRVFTSKS 1121
Qy 910 DVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVTYIMVMKCMWIDS 969
Db 1122 DVMAFGVTIWELLTFQORPHENIPAKDIPDLTIEVGLKLEQPEICSLDIYCTLLSCWHLDA 1181
Qy 970 ECRPFRELVEFSRWARDPQRFVVIQNEIDLG--PASPLDSTFYRSLLEDD---DMGDLV 1024
Db 1182 AMRPTFKLTTVFABFADPGRYLAIPGDKFRLPA-----YTSQDEKDLIRKLAPT 1234
Qy 1025 DAEYLVLPQOQFCFPDPAFGAGGMVHHRHRSSTRSGGDLTLGLEPSEERAP-----R 1078
Db 1235 DGSEAIAPDDYLOPKAAPGS-----HRTDCT-----DEMPKLRVCK 1273
Qy 1079 SPLAPSEGAGSDVFG----DLGMAAKGLQSLPHTHPSPLQRYSEDPVPLPSETDGYVA 1135
Db 1274 DPSKNKSTGDDERDSSAREVGVGNL-----LDLPVDEDDYLM 1312
Qy 1136 PLTCSPOEYVNPQVVRPQPSREGPLPAARPAATLERAKTLSFGKNGVGVKQVFAFG 1195
Db 1313 P-TCPGPNNNNNMN-----NPNQNNMAVGAAGYM-----DLIGVPV 1350
Qy 1196 AVENPEYL---TPOGGAPOPH-----PPPAFSP-AFDNLYWD 1230
```

```
Db 1351 SVDNPEYLLNAQTLGVGESPIPTQTIGIPVMGGPGTMEVKVPMGPSTSSDHEYND 1408

RESULT 12
ERBB ALV STANDARD; PRT; 634 AA.
ID ERBB ALV
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilssen T.W., Maroney P.A., Goodman R.G., Rottman F.M.,
RT Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INIT.
CC PIR; A00643; TVCLV.
CC PIR; B00643; TVFLV.
CC HSP; P11362; LFQK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.6%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 6.6e-88;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;
```

```
Qy 593 CAHYKDPFPCVARGKVPKDLSPYPIWKFPDEEGACQPCPCINCHSCVDLDDKGPAGQ 652
Db 3 CAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCTRGCCKGLEGCP--- 58
Qy 653 RASPLTSIVSAVV-GILLVWLVGVVFGILIKRQOKIRKYTRRRLQLQETELVELPTPSGA 711
```

```
Db 59 NSGKTPSIAGVVGGLLVVGLGIGLYLRR-HIVKRTLRLLQERLVEPLTPSGE 117
Qy 712 MPNQAMRILKETELRKVKVLGSGAGFTYVKGHWIPDGNVKIPVAIKVLENTSPKANK 771
Db 118 APNOAHLRLKETEFKKVKVLGSGAGFTYVKGHWIPEGEKVKIPVAIKELREATSPKANK 177
Qy 772 EILDEAYVWAGVSPVSRLLGLTSTVQLVTLMPYGLLDHVRNRLGSGDOLLNW 831
Db 178 EILDEAYVWASVNDPHVCHRLGLTSTVQLITQLMPYGLLDYIREHKNDIGSQYLLNW 237
Qy 832 CMQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 951
Db 238 CVQIAKGMNLEERLVRHDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGKVP 297
Qy 892 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQPP 357
Qy 952 ICTIDVYIMVKCWMIDSCRPRERELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYIMVKCWMIDASRPKRELIAEFSKWARDPPRYLVQIGDERHKLPSPTDSKF 417
Qy 1011 YRSLLDEDDMDGLVDAAEYLVPOQGFPCPDPAAGAGMWVHRRHSRSTSGGGDLTLGLE 1070
Db 418 YRTLMEEDEMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEEEAPRPL-----APEGAGSDVFDGDLGMAKGLQSLTPHDPSPLOYSEPTVP 1125
Db 450 -----SRTPLLSLSATSNNSATCID-----RNGQGHVPVREDVSFQRYSSDPTGN 495
Qy 1126 LPSET--DGVAPLTCSPQEVYVQDVRPQPPSPREGPLPAARPAGATLERAKTISPGK 1183
Db 496 FLEESIDDFL-----PAPEYVQ--LMPKES-----TAMVQ 526
Qy 1184 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPHPPAFSPAFONLY 1228
Db 527 NQIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTQSPKAKTVFSSPY 578
Qy 1229 WQO-----DPPE-----RGAPPSTFKGTPTAENPEYLCGLDVP 1260
Db 579 WQSGNHQINLONPDYQDFLFPNETKPNGLLKVPAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
AC F07535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
```

```
CC tyrosine phosphate.
CC DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02006; AAA42394.1; ALT_INIT.
CC EMBL; K01216; AAA42400.1; -.
CC PIR; A00644; TVYUH.
CC HSSP; P11362; LFQK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; Tyrc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
KW DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBCDD06745D609 CRC64;

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 2.1e-85;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

Qy 593 CAHYKDPFCVACPSGVKPDLSYMPIMKFPDDEGACPCPINCTHSCVDLDDKGCPEAQ 652
Db 3 CAHFDGPHCVKACPAVLGENDTL--VRKYADANAVCOLCHPNCTRGCKGPGLEGCP--- 58
Qy 653 RASPLTSIVSAVV-GILLVVLGVVFGILIKRQKIRKYTMRLQLQETELVEPLTPSGA 711
Db 59 NSGKTPSIAGVVGGLLVVGLGIGLYLRR-HIVKRTLRLLQERLVEPLTPSGE 117
Qy 712 MPNQAMRILKETELRKVKVLGSGAGFTYVKGHWIPDGNVKIPVAIKVLENTSPKANK 771
Db 118 APNOAHLRLKETEFKKVKVLGSGAGFTYVKGHWIPEGEKVKIPVAIKELREATSPKANK 177
Qy 772 EILDEAYVWAGVSPVSRLLGLTSTVQLVTLMPYGLLDHVRNRLGSGDOLLNW 831
Db 178 EILDEAYVWASVNDPHVCHRLGLTSTVQLITQLMPYGLLDYIREHKNDIGSQYLLNW 237
Qy 832 CMQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 891
Db 238 CVQIAKGMNLEERLVRHDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGKVP 297
Qy 892 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQPP 357
Qy 952 ICTIDVYIMVKCWMIDSCRPRERELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYIMVKCWMIDASRPKRELIAEFSKWARDPPRYLVQIGDERHKLPSPTDSKF 417
Qy 1011 YRSLLDEDDMDGLVDAAEYLVPOQGFPCPDPAAGAGMWVHRRHSRSTSGGGDLTLGLE 1070
```

```

Db 418 YETLMEEDMEDIVDAEVLVPHQGF-----NSPST----- 449
QY 1071 PSEEAAPRSL-----APSEGAGSDVFDGLGMAAGKQSLQSLTHDPSPLORYSEDPYV 1125
Db 450 -----SRTPLLSLSLSATSNNATNCID-----RNGQGHVPVREDSFVQRYSSDPTGN 495
QY 1126 LPSET--DGYVAPLTCSPQPEYVQPDVVPQPSREGPLPAARPAAGATLERAKTLSPK 1183
Db 496 FLEESIDQGL-----PAPEYVQ--LMPKKPSTAM----- 524
QY 1184 NGVVKDVFAF-----GGAVENPEYLTPOGGAAPQHPPPAFSPAFD 1224
Db 525 --VQNIYNFISLTAISKLPMSDRYQNSHSTAVDNPYL-----NTNQSPFLAKTVPE 574
QY 1225 NLYYVQDDPPERGAPPSTFGTPTAENPEY 1254
Db 575 SSPYIQSGNHQ-----INLDNPDY 594

RESULT 14
ERBB_AVIEU
ID ERBB_AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells.";
RL Mol. Cell. Biol. 6:1751-1759(1986).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13179; AAA42401.1; -
CC DR PIR; A25231; TVFVEB.
CC DR HSP; P11362; IFGB.
CC DR InterPro; IPR000719; Euk pkinase.
CC DR InterPro; IPR001245; Tyr_pkinase.
CC DR Pfam; PF00069; pkinase; 1.
CC DR ProDom; PD000001; Euk_pkinase; 1.
CC DR SMART; SM00219; TyrKc; 1.
CC DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;

Query Match 23.9%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 1.7e-81;

```

```

Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;
QY 593 CAHYKDPFCVACRPSGVKPDLSYMPIKFPDSEAGACQPCPINCTHSCVLDLDDKGCPAEQ 652
Db 3 CAHFIDGPHCVKACPAVLGENDTL--VMKYADANAVCOLCHPNCTRCCKGPGLEGCP--- 58
QY 653 RASPLTSIVSAVV--GILLVVVGVFGIILIKRQOKIRKYTMRRLLQETELVELPLTPSGA 711
Db 59 NGSKTPSIAAGVGVGLLCLVVGIGLYLRR--HIVRKRTLRLRLQERELVELPLTPSGE 117
QY 712 MENOQMRILKETEUKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 771
Db 118 ANQAHRLRLKETEUKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 177
QY 772 EILDEAYVMAGVSPVSRLLGICLTSTVQLVTPQLMPYGCCLDHRNRRGLSQDILLNW 831
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCCLLDYIREHKDNIGSOYLLNW 237
QY 832 CMQIAKGMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGKVP 891
Db 238 CVQIAKGMNLEERHMHVHRDLAARNVVKTPQHKITDFGLAKQLGADEKEYHAEGKVP 297
QY 892 IKWMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQPP 357
QY 952 ICTIDVIMVVKWIMIDSECRPRFRELSEFSEFMRMARDPQRFVVIQ--NEDLGASPLDSTF 1010
Db 358 ICTIDVIMVVKWMSDADSRPKFRELIAEFSEFMRMARDPQRFVVIQ--NEDLGASPLDSTF 417
QY 1011 YRSLLDEDDMDGLVDAEYLVPOQGFCCPDPAAGAGGMVHRRHRSSTRSGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1071 PSEEAAPRSL-----APSEGAGSDVFDGLGMAAGKQSLQSLTHDPSPLORYSEDPYV 1125
Db 450 -----SRTPLLSLSLSATSNNATNCIDRNGG-----H----- 476
QY 1126 LPSETDGYVAPLTCSPQPEYVQPDVVPQPSREGPLPAARPAAGAT--LERAKTLSPGKN 1184
Db 477 -PVREDGFL-----PAPEYVQ--LMPKKPSTAMVQNIYVLSITASKLPIDSRYN 527
QY 1185 GVVKDVFAPGGAVENPEYL 1203
Db 528 -----SHSTAVDNPYL 539

RESULT 15
EGFR_CHICK
ID EGFR_CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE (Fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax J., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha ";
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND

```

CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20386; AAA48760.1; -;
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; PARTIAL.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; PARTIAL.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 667 POTENTIAL.
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 703 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Query Match 22.1%; Score 1510; DB 1; Length 703;
Best Local Similarity 42.2%; Pred. No. 7.3e-75;

Matches 304; Conservative 114; Mismatches 253; Indels 50; Gaps 14;
QY 8 RMGLLLALLPPGAA-----STQVCTGTDMKRLRASPETHLDMLRHLRYOGCOVVOGNLE 61
DB 13 RGAALVLLLLGLVALCSAVEEKVCOGTNNKLTQLGHVEDHFTSLQRMNNYCNVLSNLE 72
QY 62 LTYLPTNASLFLQDIQEVQGVVLIATNOVROVPLQRLIRVGTQLFEDNYALAVLDNGD 121
DB 73 IYVEHNRDLTLFKTIQEVAGIVLALANMVDVPLENLIIRGNVLYDNSFALAVLSNYH 132
QY 122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNQAL 181
DB 133 -MNTQ-----GLRELPMKRLSEILNGGVKISNNPKLCNNMTVLWMDIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPCSPCKGRCWSESSEDCSLTRTYCAGCA-RCKGPLFPNNFTVS 239
DB 183 TVLDFASNUSSPCKHPNCTEDHCWAGEQNCOTLTKVCAQQCGRCKGVKPSD----- 237
QY 240 FWLVRPKVSASHLE-----KHSDCLACLHFNHSGICELHCPALVTVYNTDTFESMNP 291
DB 238 -----CCHNQCAAGCTGPRSDCLACRFRDDATCKDTCPLVLYNPTTYQMDVNP 288
QY 292 EGRYTFGASCVTACPNYLSLTVGSCTLVCPLNHQEVTAADGTQRCCKSPCARCYGL 351
DB 289 EGKYSFGATCVRECPHNYVVDHGSCVRSCTDITYEV-BENGVRKCKCDGLCSKVCGI 347
QY 352 GMEHLREVRAVTSANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVFTLEE 411
DB 348 GIGELKGLISINATNIDSFNCTKINGDVSIILPVAFGLDAFTKTLPLDPKKLDVFRVKE 407
QY 412 ITGYLIYISAWPDSLPLSVFQNLQVIRGIRLHNGAYSLTLQGLIGISWLGSLRSLRELGSGL 471
DB 408 ISGFLLIQAWPDNATDLYAFENLEIRGTQKHQGYSLAVVNLKIQSLGLRSLKEISDGD 467
QY 472 ALIHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHGWGP 531
DB 468 IAIMKNKLCYADTMNRSLFATQSOKTKIIQNRNKNDCDADRHRVCDPLCSDVGCWGP 527
QY 532 TQVNCVSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECOPEG---SVTCFGPEAD 588
DB 528 FHCFSRFRSROKECVKQCNILQGEPEFERDSKLPCHSECLVQNSTAYNTTCSGGPD 587
QY 589 QCVACAHYKDPFPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGC 648
DB 588 HCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCQLCHPNCIRGCKGPGLEG 646
QY 649 PAEQRASPLTSTVSADV-GILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLT 707
DB 647 P---NGSKTPSIAAGVVGGLCLLVVVGILGILYLRRL-HIVRKRTLRLRLQERLVEPLT 702
QY 708 P 708
DB 703 P 703

Search completed: July 22, 2003, 08:45:17
Job time : 21.2793 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-210-224-14
Perfect score: 6827
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacterioph.*
- 17: sp_archae.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 6173 | 90.4 | 1259 | 6 | O18735 |
| 2 | 3057 | 44.8 | 1209 | 11 | Q9GX70 |
| 3 | 3028 | 44.4 | 1210 | 11 | Q9EP98 |
| 4 | 2652 | 38.8 | 1165 | 13 | Q9VH40 |
| 5 | 2631.5 | 38.5 | 1137 | 13 | Q9W6F6 |
| 6 | 2238 | 32.8 | 1328 | 13 | Q9W754 |
| 7 | 1953.5 | 28.6 | 1433 | 5 | Q9BIH9 |
| 8 | 1752.5 | 25.7 | 419 | 4 | Q9UK79 |
| 9 | 1739 | 25.5 | 367 | 11 | Q8R2X1 |
| 10 | 1720 | 25.2 | 729 | 15 | Q8E712 |
| 11 | 1718 | 25.2 | 567 | 15 | Q8E714 |
| 12 | 1697.5 | 24.9 | 412 | 4 | Q8WYV0 |
| 13 | 1655.5 | 24.2 | 962 | 15 | Q64895 |
| 14 | 1645 | 24.1 | 545 | 15 | Q85468 |
| 15 | 1419.5 | 20.8 | 655 | 11 | Q9WVF5 |
| 16 | 1403.5 | 20.6 | 643 | 11 | Q9ERV6 |

| | | | | | |
|----|--------|------|------|----|--------|
| 17 | 1211 | 17.7 | 1193 | 5 | Q9Y1X8 |
| 18 | 1157.5 | 17.0 | 1368 | 5 | Q23821 |
| 19 | 1142 | 16.7 | 1717 | 5 | Q26566 |
| 20 | 1041 | 15.2 | 527 | 13 | Q90836 |
| 21 | 973.5 | 14.3 | 599 | 13 | Q9PSH2 |
| 22 | 914.5 | 13.4 | 478 | 11 | Q9ESE0 |
| 23 | 906 | 13.3 | 165 | 4 | Q14256 |
| 24 | 887 | 13.0 | 176 | 11 | Q923V5 |
| 25 | 806.5 | 11.8 | 346 | 13 | P11776 |
| 26 | 778 | 11.4 | 435 | 5 | Q8SZW1 |
| 27 | 754.5 | 11.1 | 311 | 13 | Q9P162 |
| 28 | 707.5 | 10.4 | 1362 | 13 | Q9PVZ4 |
| 29 | 678 | 9.9 | 1671 | 5 | Q9NJV5 |
| 30 | 658.5 | 9.6 | 1368 | 13 | Q9UW85 |
| 31 | 638 | 9.3 | 1418 | 13 | Q93457 |
| 32 | 636 | 9.3 | 331 | 4 | Q9BUD7 |
| 33 | 625 | 9.2 | 1358 | 13 | Q73798 |
| 34 | 623.5 | 9.1 | 1472 | 5 | Q9U5A8 |
| 35 | 620.5 | 9.1 | 1369 | 13 | Q8UW86 |
| 36 | 604 | 8.8 | 1412 | 13 | Q8UW84 |
| 37 | 601 | 8.8 | 149 | 6 | Q9BG66 |
| 38 | 596 | 8.7 | 987 | 11 | Q91YMO |
| 39 | 596 | 8.7 | 1371 | 11 | Q9QVW4 |
| 40 | 593 | 8.7 | 987 | 11 | Q99MR2 |
| 41 | 588.5 | 8.6 | 1418 | 13 | Q8UW83 |
| 42 | 587.5 | 8.6 | 1036 | 4 | Q07912 |
| 43 | 587.5 | 8.6 | 1055 | 11 | Q54967 |
| 44 | 586.5 | 8.6 | 1091 | 4 | Q9DMQ4 |
| 45 | 584.5 | 8.6 | 2144 | 5 | Q9VD94 |

ALIGNMENTS

RESULT 1

| | | | |
|----------|---|------|----------|
| O18735 | PRELIMINARY; | PRT; | 1259 AA. |
| ID | O18735; | | |
| AC | O18735; | | |
| DT | 01-JAN-1998 (Tremblrel. 05, Created) | | |
| DT | 01-JAN-1998 (Tremblrel. 05, Last sequence update) | | |
| DT | 01-JUN-2002 (Tremblrel. 21, Last annotation update) | | |
| DE | Erbb-2. | | |
| (| Canis familiaris (Dog). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis. | | |
| OX | NCBI_TaxID=9615; | | |
| RN | [1]_TaxID=9615; | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | Yokota H.; | | |
| RT | "cDNA cloning of erbB-2 from canine mammary gland." | | |
| RL | Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL; AB008451; BAA23127.1; | | |
| DR | HSSP; F11362; 1FGK. | | |
| DR | InterPro; IPR002048; EF-hand. | | |
| DR | InterPro; IPR000494; EGFR_L_domain. | | |
| DR | InterPro; IPR000719; Euk_pkinase. | | |
| DR | InterPro; IPR002174; Furin-like. | | |
| DR | InterPro; IPR001245; Tyr_pkinase. | | |
| DR | InterPro; IPR004019; YLP_motif. | | |
| DR | Pfam; PF00757; Furin-like; 1. | | |
| DR | Pfam; PF00069; pkinase; 1. | | |
| DR | Pfam; PF01030; Recep_L_domain; 2. | | |
| DR | Pfam; PF02757; YLP; 2. | | |
| DR | ProDom; PD000001; Euk_pkinase; 1. | | |
| DR | SMART; SM00261; FU; 3. | | |
| DR | SMART; SM00219; TyrKc; 1. | | |
| DR | PROSITE; PS00018; EF_HAND; UNKNOWN 1. | | |
| DR | PROSITE; PS00107; PROTEIN_KINASE_ATP; 1. | | |
| DR | PROSITE; PS00111; PROTEIN_KINASE_DOM; 1. | | |
| DR | PROSITE; PS00109; PROTEIN_KINASE_TYR; 1. | | |
| KW | ATP-binding; Transferase; Tyrosine-protein kinase. | | |
| SEQUENCE | 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64; | | |

| | |
|--------|-------------|
| Q9Y1X8 | ephydatia f |
| Q23821 | caenorhabdi |
| Q26566 | schistosoma |
| Q90836 | gallus gall |
| Q9PSH2 | gallus gall |
| Q9ESE0 | rattus norv |
| Q14256 | homo sapien |
| Q923V5 | rattus norv |
| P11776 | xiphophorus |
| Q8SZW1 | drosophila |
| Q9P162 | xiphophorus |
| Q9PVZ4 | xenopus lae |
| Q9NJV5 | biomphalari |
| Q8UW85 | paralichthy |
| Q93457 | scophthalmu |
| Q9BUD7 | homo sapien |
| Q73798 | xenopus lae |
| Q9U5A8 | bombyx mori |
| Q8UW86 | paralichthy |
| Q8UW84 | paralichthy |
| Q9BG66 | oryctolagus |
| Q91YMO | mus musculu |
| Q9QVW4 | rattus sp. |
| Q99MR2 | mus musculu |
| Q8UW83 | paralichthy |
| Q07912 | homo sapien |
| Q54967 | mus musculu |
| Q9DMQ4 | homo sapien |
| Q9VD94 | drosophila |

| | | |
|----------------------------|------|--|
| Query Match | | 90.4%; Score 6173; DB 6; Length 1259; |
| Best Local Similarity | | 90.0%; Pred. No. 0; |
| Matches 1146; Conservative | | 40; Mismatches 60; Indels 28; Gaps 4; |
| QY | 1 | MELAAACRGLLLALLPPGAASQTQCTGDMKRLPASBETHDMLRHLYQGQVQGNL 60 |
| DB | 1 | MELAAACRGLLLALLPSGAAGTQVCTGDMKRLPASBETHDMLRHLYQGQVQGNL 60 |
| QY | 61 | ELTYLPTNASLSFLQDIEQVQVVLTAHQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120 |
| DB | 61 | ELTYLPANASLSFLQDIEQVQVVLTAHQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120 |
| QY | 121 | DPUNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITILMKDIFPHKNQLA 180 |
| DB | 121 | DPLEGGIPAPGAAGGLRELQRLSLTEILKGGVLIQRNPOLCHODITILMKDVPHKNQLA 180 |
| QY | 181 | LTLIDTNRSRACHPCSPMKSGSCWGESSEDCSLTRTVCCAGCARCKGPLPNNNTVSF 240 |
| DB | 181 | LTLIDTNRFSACPPCSPACKDAHCWASSGDCOSLTRTVCCAGCARCKGPQ----- 232 |
| QY | 241 | WLRVPKVSASHLE-----KHSDCIACILFHNHSGICELHCPALVTYNTDTPESMPNPE 292 |
| DB | 233 | -----TDCHECOAGCTGPKHSDCLACILFHNHSGICELHCPALVTYNTDTPESMPNPE 286 |
| QY | 293 | GRYTFGASCVTACPYNYLSTDVSGCTLVCLPLHNOEVTABDGTQRCCKSKPCARVCYGLG 352 |
| DB | 287 | GRYTFGASCVTSCPYNYLSTDVSGCTLVCLPLNNOEVTABDGTQRCCKSKPCARVCYGLG 346 |
| QY | 353 | MEHLREVRVANTSANIOEFAGCKIFGSLAFLPESFDGDPASNTAPLOPOLQVFTLEI 412 |
| DB | 347 | MEHLREVRVANTSANIOEFAGCKIFGSLAFLPESFDGDPASNTAPLOPOLRVEALEEI 406 |
| QY | 413 | TGYLYISAWPDSLPDLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLA 472 |
| DB | 407 | TGYLYISAWPDSLPLNSVFQNLRVIRGRVLHDGAYSILTQGLGISWGLRSLRELGSGLA 466 |
| QY | 473 | LIHNTHLCFVHTVPMDQLFRNPQHALLTANRPEDECVEGGLACHOLCARGHCWPGPT 532 |
| DB | 467 | LIHNRNARLCFVHTVPMDQLFRNPQHALLHSANRPEBECVGEGLACYPC-CAHGHGCPGPT 525 |
| QY | 533 | QCNCVSQFURGQECVECRVLQGLPREYVNAHCLCHPECOPONGSVTCFGEADQOVA 592 |
| DB | 526 | QCNCVSQFURGQECVECRVLQGLPREYVKDRYCLPCHSECQPNQSGVTCFGEADQOVA 585 |
| QY | 593 | CAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEBEGACQPCINCTHSCVDLDDKGCPAEQ 652 |
| DB | 586 | CAHYKDPPFCVARCPSGVKPDLSFMPWKFADESGTCQPCINCTHSCADLDEKGCPAEQ 645 |
| QY | 653 | RASPLTISVAVVGIILLVVLGVVFGILIKRROQKIRKYTMRLLOQETELVEPLTPSGAM 712 |
| DB | 646 | RASPVTSIIAAVVGIILLVAVVGLVGLILIKRRRQKIRKYTMRLLOQETELVEPLTPSGAM 705 |
| QY | 713 | PHQAQWRIILKTELKRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 772 |
| DB | 706 | PHQAQWRIILKTELKRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 765 |
| QY | 773 | ILDEAYMAGVSPYVSRLLGICLTSTVQLVTLQMPYGLLDHVRENRLGASQDLLNMC 832 |
| DB | 766 | ILDEAYMAGVSPYVSRLLGICLTSTVQLVTLQMPYGLLDHVREHRLGASQDLLNMC 825 |
| QY | 833 | MQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 892 |
| DB | 826 | VQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 885 |
| QY | 893 | KWMALESILRRRFTHOSDWSVGVTVWELMTGAKPYDGIIPAREIDDLLEKGERLPQPI 952 |
| DB | 886 | KWMALESIPPRFTHOSDWSVGVTVWELMTGAKPYDGIIPAREIDDLLEKGERLPQPI 945 |
| QY | 953 | CTIDVYIMVKCWMIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEEDLGPASPLDSTFYR 1012 |
| DB | 946 | CTIDVYIMVKCWMIDSECRPRFRELVAEFSRMARDPQRFVVIQNEEDLGPASPLDSTFYR 1005 |
| QY | 1013 | SILLEDDMDGLVDAEEYLVPOQFFCDDPAPGAGMWHHRHSSSTRSGGDLTLGLEPS 1072 |

Query Match 44.8%; Score 3057; DB 11; Length 1209;
Best Local Similarity 48.8%; Pred. No. 2.7e-22;

| Matches | 530; | Conservative | 168; | Mismatches | 357; | Indels | 136; | Gaps | 28; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------|------|--------------|-----------|------------------|------------|-----------|-------------|----------|--------------|----------------|--------------|--------|----------|-----------|---------|-------|-------|------|----------|-------------|-----|----|----|----|---|---|---|---------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-------|-------|-------|---|-------|---|-------|-------|
| Qy | 3 | LAALCRG | LLALLPPGA | -ASTOYCTGTGDMKLR | LRLPASPETH | DLMLRHL | YQCCVQVGNLE | 61 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 15 | LAALCAAG | ----- | -GALEEKV | CCQGT | SNRLTQ | LGTGFEDH | FLSQRMFN | CEVVLGNLE 66 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Qy | 62 | LTYYLPT | NASLSD | FLQDIOE | VQVYVLI | IAHNVQ | RVPLRLR | IVRG | TQOLFEDNY | VALVDNGD 121 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 67 | ITYVQ | RNYDL | SLFKTIOE | VAGYVLI | ALTNTVERI | PLENLQI | IRGNAL | YENTY | VALAVLSN-- 124 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Qy | 122 | PLNNT | TPVTGAS | PGGLRE | LQSLR | LEILK | GGVLR | LORN | POLCYQ | DTILKWKDIFHKN | NOLAL 181 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 125 | ----- | -YGNK | TGLRE | LPWRN | QELI | IGAVRS | NNPIL | CNMETI | QWRDIV | -ODVFLSN 175 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Qy | 182 | TLID | TNRS | -RACHP | CSBMK | SGRCW | GESSD | CQSL | TRTV | CAGGCA | -RCKG | PLPENN | FTVS 239 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 176 | MSMD | VQRHL | TGCPK | CDP | CPNGS | CMW | GEENC | QKLT | KII | CAQCS | RRCR | GRSPD | ----- 230 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Qy | 240 | PWL | RVPK | VASHLE | ----- | -KHSD | CLAC | LHF | NHSGI | CE | LHC | PAL | VTYNT | DTFES | MNP 291 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 231 | ----- | -CCHNQ | CAAG | CTG | PRES | DC | LVCH | RFDE | ATCK | DTCP | PLML | NYPT | TYQMD | VNP 281 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Qy | 292 | EGR | YTFG | ASCV | TAC | PNY | LS | TDV | GS | CT | LV | CP | LN | QEV | TAE | DGTOR | CEKSK | PCAR | VCYL 351 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 282 | EKG | YSF | GAT | CVK | KPR | NYV | TDH | GS | CV | RAC | GP | DY | YEV | -BED | GS | VSKK | CKD | GP | CKVCNGI 340 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Qy | 352 | GME | HLR | EV | RA | VS | ANI | QEF | AG | CKI | FG | S | LA | FL | PES | FD | GP | DS | PA | SN | TAP | LQ | PE | LO | V | F | T | LEE 411 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 341 | GIG | E | K | D | T | L | S | I | N | A | T | I | K | H | F | Y | C | T | A | I | S | G | D | L | H | L | P | V | A | F | K | G | D | S | F | T | R | P | P | L | D | P | R | E | L | I | K | T | V | K | E 400 | | | | | | |
| Qy | 412 | I | T | G | V | L | T | S | A | M | P | D | S | L | P | D | S | V | F | O | N | L | O | V | I | R | G | R | I | L | H | N | G | A | S | L | T | I | O | G | I | S | W | L | R | S | L | R | E | L | G | S | L 471 | | | | | |
| Db | 401 | I | T | G | F | L | I | Q | A | W | P | N | T | D | L | H | A | F | E | N | L | E | I | R | G | R | T | Q | H | Q | F | S | L | V | V | G | L | N | I | T | S | L | R | S | L | K | E | I | S | D | G 460 | | | | | | | |
| Qy | 472 | A | L | I | H | N | T | L | C | F | H | V | T | P | M | D | L | F | R | N | P | H | O | A | L | L | T | A | N | R | P | E | D | C | V | G | E | G | L | A | C | H | O | L | C | A | R | G | H | C | W | P | G 531 | | | | | |
| Db | 461 | V | I | S | G | R | N | L | C | V | A | N | T | I | N | N | K | L | F | G | T | P | N | Q | T | K | I | M | N | N | R | A | E | K | D | C | K | A | T | N | H | V | C | N | P | L | C | S | S | E | G | C | W | P | E 520 | | | |
| Qy | 532 | T | Q | C | V | N | C | S | O | F | L | R | G | Q | E | C | V | E | C | R | V | L | Q | L | P | R | E | Y | N | A | R | H | C | L | P | C | H | P | E | C | O | P | Q | N | G | S | V | T | C | F | G | P | E | A | D | O | C | V 591 |
| Db | 521 | T | D | C | V | S | C | N | V | S | R | G | E | C | V | D | K | N | I | L | E | G | E | P | R | E | F | V | E | N | S | E | C | I | O | C | H | P | E | C | L | P | O | T | M | N | I | T | C | T | G | R | G | P | N | C | I 580 | |
| Qy | 592 | A | C | A | H | Y | K | D | P | P | C | V | A | R | C | P | S | G | V | P | D | L | S | Y | M | P | I | W | K | P | P | E | E | G | A | C | Q | P | C | I | N | T | H | S | C | V | D | L | D | D | K | G | P | A | E 651 | | | |
| Db | 581 | K | C | A | H | Y | D | G | P | C | V | K | T | C | P | S | G | I | M | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 3
Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7TAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF275366; AAG28045.1; -.
EMBL; AF275364; AAG28045.1; JOINED.
EMBL; AF275365; AAG28045.1; JOINED.
EMBL; AF275367; AAG24386.1; -.
HSP; P11362; IFGK.
MGD; MGI:95294; Egfr.
InterPro; IPR000345; CytC heme bind.
InterPro; IPR000494; EGFR_L domain.
InterPro; IPR000719; Euk_pkinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR002290; Ser_thr_pkinase.
InterPro; IPR001245; Tyr_pkinase.
Pfam; PF00757; Furin-like; 1.
Pfam; PF00069; pkinase; 1.
Pfam; PF01030; Recep_L domain; 2.
PRINTS; PR00109; TYRKINASE.
ProDom; PD000001; Euk_pkinase; 1.
SMART; SM00261; FU; 5.
SMART; SM00220; S_TKC; 1.
SMART; SM00219; TYRK; 1.

```

```
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; Receptor; Transferase.
SQ SEQUENCE 1210 AA; 134940 MW; 62CD021C9DE32E18 CRC64;

Query Match 44.4%; Score 3028; DB 11; Length 1210;
Best Local Similarity 48.3%; Pred. No. 4.2e-220;
Matches 621; Conservative 170; Mismatches 363; Indels 132; Gaps 25;

QY 11 LLALPPGAA--STVCTGTDMKRLPASPETHLMDLRHLXGCGVQVGNLELTPLPTN 68
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 ASLSFLQDIQEVGYVLIAHNOVROVPLORLVRGTQLPEDNYALAVLQNGDPLNNTTP 128
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 VTGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYQDTILWKDI----FKNNQLALTLI 184
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 DTNRSRACHPCSPMKGSRGWSGESSDDCQLTRTVGAGCA--RCKGPLPNNFTVFWLR 243
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 VPKVSASHLE-----KHSDCLACLFHNSHSGICELHCPALVYNTDTFESMPNPGRY 295
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 296 TFGASCVTACPNYLSLTDVSGCTLVCPLNHNOETAEADGTQRCCKSKPCARVCYGLGMEH 355
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 356 LREVRVAVTSANIOEPAGCKIFGSLAPLPSFGDPSASNTAPLOPSOLQVFTELEITGY 415
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 416 LYSISAMPDSLPLDSVFQNLQVIRGLIHNGAYSLTGLGIGLWGLRSRLSGSLALIH 475
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 476 HNTLFCVHTVPMDQLFRNPHQALHTANRPEDECYGEGLACHQLCARGHCWGGPTQCV 535
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 536 NCSQFLRGQECVECEVQLQGLPREYVNAHCLCHPECPQNGSVTCFGEADQCACAH 595
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 596 YKDPFPCVARCPGVPDLISYMPIKFPDBEGACQPCPINCSTHSCVDLDDKGCAPQORAS 655
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 656 PLTSIVSAVVGILLVVLVGVVFGI--LIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPN 714
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 715 QAQMRLKETELRKVKVLGSGAGTYVKGWIIPDGENVKIPVAIKVLRENTSPKAKEIL 774
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 775 DEAYVMAGVSPVSVRLGLICTSTVOLVTQLMPYGCLLDHDVRENRLGSGQDLLNWCWQ 834
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 835 IAKGMSYLEDLVRLVHRDLAARNVLVKS PNHVKTITDFGLARLLIDIDETEHADGKGVPIKW 894
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 895 MALESILRRRFTHQSDVWSYGVTVWELMTFGSKPYDGI PASDISSILEKGERLPQDPIC 942
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 955 IDVYIMVWKCMWIDSECRFRRELVSERFARMARDPQRFVVIQ--NEDLGPASPLDSTFYRS 1013
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 943 IDVYIMVWKCMWIDADSRKFRLEILILEFSKMDARQRYLVIIQDGRMHLPSDTSNFYRA 1002
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1014 LLEDDDMGDLVDAEYLYVPOQGFPCPDPAAGAGGMVHHRSSSTRSGGDLTLGLEPSE 1073
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1074 EBAPRSPLAPSEGAGSDVFDGLGMAAKGLOSLPHDPSPLQRYSDPTVPLPSET--D 1131
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1032 ---SRTPLSSLSATSN---NSTVACINRNGSCRVKEDAFLOQYSSDPTCAVTEIDID 1084
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1132 GYVAPLTCSPQEVYVNPQVVRPPOPPSPREGPLPAARPAAGATLERAKTLSPKNGVVKDVF 1191
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1192 AFGGAVENPEYL--TPQGAAPQPHPPAFSPAFDNLYYWDQ-----DP-----P 1234
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1235 ERGAPPSTFKGTPTAENPEYLGLDVP 1260
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1179 KETKPNIGIEFG-PTAENAEYLRVAPP 1203
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
Q9YH40 PRELIMINARY; PRT; 1165 AA.
AC Q9YH40;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase proto-oncogene.
GN XMRK.
OS Xiphophorus xiphidium.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8086;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RX MEDLINE=98241172; PubMed=9582016;
RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RA Altschmid J., Scharl M.;
RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT overexpression and mutational alterations.";
RL Oncogene 16:1681-1690(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RA Scharl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53471; AAD10500.2; --
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; Cyt_c_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
```


| | | | |
|---|--|--|------|
| Db | 997 | PVEENPFVRRKNGDLQAVDNPEYHN-----APNGQPKAEDEVYVNEPLYLNTFANTLE | 104 |
| Qy | 1199 | NPEYLTPOGGAAPQHPPPAFDPNLYYNDODDPERGA--PPSTFKGTPT----- | 1248 |
| Db | 1050 | NAEYL-----KNNLPEKAKKAPDNDPMYHNSLPPRSTLQHPDYLOEYSTKYFYKQNG | 1101 |
| Qy | 1249 | -----AENPEYL | 1255 |
| Db | 1102 | RIRPIVAENPEYL | 1114 |
| RESULT 6 | | | |
| ID | P79754 | PRELIMINARY; PRT: 1328 AA. | |
| AC | P79754; | | |
| DT | 01-MAY-1997 | (TRENBLrel. 03, Created) | |
| DT | 01-MAY-1997 | (TRENBLrel. 03, Last sequence update) | |
| DT | 01-JUN-2002 | (TRENBLrel. 21, Last annotation update) | |
| DE | Erbb3. | | |
| OS | Fugu rubripes (Japanese pufferfish) (Takifugu rubripes). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; | | |
| OC | Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; | | |
| OC | Tetraodontidae; Takifugu. | | |
| OX | NCBI_TaxID=31033; | | |
| RP | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | MEDLINE=99177347; PubMed=10077531; | | |
| RA | Gelliner K., Brenner S.; | | |
| RT | "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu | | |
| RT | rubripes."; | | |
| RL | Genome Res. 9:251-258(1999). | | |
| DR | EMBL; AF056116; AAC34391.1; -- | | |
| DR | HSSP; P11362; 1FGK. | | |
| DR | InterPro; IPR000494; EGFR_L_domain. | | |
| DR | InterPro; IPR000719; Euk_pkinase. | | |
| DR | InterPro; IPR002174; Furin-like. | | |
| DR | Pfam; PF00757; Furin-like; 1. | | |
| DR | Pfam; PF00069; pkinase; 1. | | |
| DR | Pfam; PF01030; Recep_L_domain; 2. | | |
| DR | ProDom; PD000001; Euk_pkinase; 1. | | |
| DR | SMART; SM00261; FU; 3. | | |
| DR | SMART; SM00219; TyrKc; 1. | | |
| DR | PROSITE; PS00107; PROTEIN_KINASE_ATP; 1. | | |
| DR | PROSITE; PS50011; PROTEIN_KINASE_DOM; 1. | | |
| KW | ATP-binding; Transferase. | | |
| SQ | SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64; | | |
| Query Match 32.8%; Score 2238; DB 13; Length 1328; | | | |
| Best Local Similarity 39.3%; Pred. No. 3, 1e-160; | | | |
| Matches 515; Conservative 153; Mismatches 411; Indels 230; Gaps 32; | | | |
| Qy | 9 | WGLLLALLPP--GAASTQ-----VCTGTDMLRLPASPEETHLMLRHLRYOGCQVVOGNLEL | 62 |
| Db | 4 | WRLILMCVASLRAASSQTQEAVCPTQGLSSTGSQENQVNLNDRYKGCBIIGNLEI | 63 |
| Qy | 63 | TYLPNASSLFLDIOEQVGVYLIHNVQVRQVLPQRLRIVRGTQLFEDNYALAVLDNGDP | 122 |
| Db | 64 | TQIESNWDPSFLKTIETREVTGYVLIAMNHFOEIPQLQRLVIRGNSLYERRFALSVELN--- | 120 |
| Qy | 123 | LNNTPTVTVGASPGGLRELQRLSLTEILAGGVLIQNPOLCYQDITLWKDIFHKNQALTL | 182 |
| Db | 121 | ----YPKDG--PSGLNQGLMLNLTLEILDGGVOIINKYLRYGFPWYWRDII--RNNDAPIE | 173 |
| Qy | 183 | LIDTNRSRACHPCSPMKSGRCWGESSEDCQSILRTVCAGGC----- | 224 |
| Db | 174 | IQFNGERGVCV--KSC-GNYCWGPGKQCCQLITKTVCAQCNDRCFCGTSPRDCCHIECA | 229 |
| Qy | 225 | ARCKGGLPFPNFTVSWLVRVPKVSASHLEKHSKSDCLACLHFNHSGICELHCPALVYNTDT | 284 |
| Db | 230 | AGCKGGL-----DTDCFACRLFNDSGACVPCQPTLIYNKQT | 266 |

| | | | |
|-----------|----------------------------|---|------|
| Qy | 285 | FESWPNBGRYTFGASCVTACPNYVLSTDVSGCTLVPLHNOEYTAEDGTOR--CEKSGKP | 343 |
| Db | 267 | FQMETNPNAKYQSGICVCSQPTHFV--VDGSSCVCPDPDKMEV--BRGSQRQCELSG | 323 |
| Qy | 344 | CARVCYGLGMEHLREVRVAVTSANTOEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEOL | 403 |
| Db | 324 | CPKVCETGAB--QROTVDSSNDISFNCTKIOGSLHFLVTGILGDDFKNVPPLDAAKL | 380 |
| Qy | 404 | QVFFETLEEITGLYISAMPDPLDLSVFQNIQVIRGRIHNGAYSLTLOGLIGISWLGJRS | 463 |
| Db | 381 | EVFRTVREITDLINQSWPKELNDLSVFSSLTTOGRSLPKRFSLMVMRIPTLTSGLRS | 440 |
| Qy | 464 | LRELGSGLALIHNNTHLJCFVHTVPWDQLFRNPH--QALLHTANRBEDECVBEGLAHOLCA | 522 |
| Db | 441 | LREISDGSVYISONAHLCYHTVNTQLFGRSFRVANSLSNRPMACEVADGRVCDPLCS | 500 |
| Qy | 523 | RGHCMGPGTQVNCSPQFLRQECVBECEVLQGLPREYVNAH--CLPCHPECOQNGSVT | 581 |
| Db | 501 | DSCGCMGPGPDQCLSCRNYSRHGTCVAGCHFNSGIPREFAGLNGVCVACHPECKPOTGAS | 560 |
| Qy | 582 | CFGPEADQCVACAHYKDPFFCVARCPGKVPDLSYMPIWKPEPDEGACQPCINCTHSCV | 641 |
| Db | 561 | CTGCADECMACTKFERDGPYCMSCSPAGW--DGEKGLIFKFPNREHCEPCNQCTGCS | 619 |
| Qy | 642 | DLDDKGPAPAEQASPLTSIVSAVVGILLVVLGVVF-----GILIKRROQKIRKYTMR | 694 |
| Db | 620 | GPGLNDC--LEAARLTSSGOITGIALGVPAGLIFCLVLFGLMLYHRGLAIRKRAMR | 676 |
| Qy | 695 | RLLOETELVEPLTSPGAMPNQAQRILKETELRKVKVLGSGAFGTVYKGIWIDGENVKI | 754 |
| Db | 677 | RYLESGESFELGP--GEKGTKVHARILKPSDLRKIKPLGSGVFGTYSKGFWIPEGETVKI | 735 |
| Qy | 755 | PVAJTKVLRENTSPRANKBEILDEAYVMAGVGSPPYVSRLLGICLTSTVOLVTOLMPYGCILD | 814 |
| Db | 736 | PVAJTKQDSSGRQTFEITDHLISMGLSDHPYIVRLGICPCGTCLQVLVTLQSHGSLLE | 795 |
| Qy | 815 | HVRENRRGLGSQDLLNMCWQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLAR | 874 |
| Db | 796 | HIRQKTSLDPÖRLNMCVQIAKGWYILEEHRVHVKNLAARNILLKNDYQVOISDYGVAD | 855 |
| Qy | 875 | LLDIDETEHADGGKVPKIMWALESIILRRRTHOSDWSYGVTVWELMTFGAKPYDGIPA | 934 |
| Db | 856 | LLYPDDKKYVYSETKPIKMWALESIILFRRYTHOSDWSYGVTVWEMMSFGAEPYASVQP | 915 |
| Qy | 935 | REIPDLLEKGBRLPOPPICTIDVYIMVWKMWIDSECRPRELVSFEFSRMAPDQRFVY | 994 |
| Db | 916 | QEVSVLEKGBRLSQPAICTIDVYMWVKMWIDENRPTFKELASOFTNRADPPRYLV | 975 |
| Qy | 995 | IQNEDLGPASPLDSTFYRSLLEDMDGLVDABEYLVPOQGFPCDPAPGAGMVHHRH | 1054 |
| Db | 976 | IRMEG-----EDSGMGEFL-----R | 990 |
| Qy | 1055 | SSSTRSGGDLTGLPSEEBEAPRSLAPSEGAGSDVFDGLMG---AAKGLQSLPTH | 1111 |
| Db | 991 | RGSR--GLLEADLEDEEE-----GLGDRFATPSLQSPSPWS | 1056 |
| Qy | 1112 | PSPLQ-----RYSEDPTVLPSETDCYVAPLTCSPQ--EYVNO----- | 1148 |
| Db | 1027 | TSPSQINSYMWYNTQLRVD-----FVSGQGHIGYLPMSPPVDITIRQLMYQSRLSVR | 1080 |
| Qy | 1149 | --PDVRPOPSPREGPL--PAARPAGATLERAKTLPQKNGVWVDVAFGGAVENPEVLT | 1204 |
| Db | 1081 | TLPDRSAFRSSRAELCEDGACAGIFVR-----FSGERN----- | 1118 |
| Qy | 1205 | POGGAOPHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPE | 1253 |
| Db | 1119 | POGG-----QQRKLSTASSPSFSFKTAADEDEE | 1146 |
| RESULT 7 | | | |
| Q9BIH9 | PRELIMINARY; PRT; 1433 AA. | | |
| ID Q9BIH9 | | | |
| AC Q9BIH9 | | | |

RESULT 7
Q9BIH9
ID Q9B
AC Q9B

| | |
|-----|--|
| DT | 01-JUN-2001 (TRENBLrel. 17, Created) |
| DT | 01-JUN-2001 (TRENBLrel. 17, Last sequence update) |
| DT | 01-JUN-2002 (TRENBLrel. 21, Last annotation update) |
| DE | Putative epidermal growth factor receptor (Fragment). |
| GN | EGFR. |
| CS | Anopheles gambiae (African malaria mosquito). |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; |
| OC | Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; |
| OC | Anopheles. |
| OX | NCBI_TaxID=7165; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RC | STRAIN=SUA; |
| RA | Lycett G.J.; |
| RT | "Cloning, expression and localisation of the Anopheles gambiae |
| RT | epidermal growth factor receptor"; |
| RL | Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases. |
| DR | EMBL; AJ301655; CAC3S008.1; - |
| DR | HSSP; P11362; IFGK. |
| I.S | InterPro; IPR000345; CyC heme bind. |
| DR | InterPro; IPR000494; EGFR_L domain. |
| DR | InterPro; IPR000719; Euk_pkinase. |
| DR | InterPro; IPR002174; Furin-like. |
| DR | InterPro; IPR002290; Ser thr_pkinase. |
| DR | Pfam; PF00757; Furin-like; 1. |
| DR | Pfam; PF00069; kinase; 1. |
| DR | Pfam; PF01030; Recep_L domain; 2. |
| DR | PRINTS; PR00109; TYRKINASE. |
| DR | ProDom; PD000001; Euk_pkinase; 1. |
| DR | SMART; SM00261; FU; 7. |
| DR | SMART; SM00220; S_TKc; 1. |
| DR | SMART; SM00219; TyrKc; 1. |
| DR | PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_4. |
| DR | PROSITE; PS00107; PROTEIN_KINASE_ATP; 1. |
| DR | PROSITE; PS0011; PROTEIN_KINASE_DOM; 1. |
| DR | PROSITE; PS00109; PROTEIN_KINASE_TYR; 1. |
| KW | Receptor. |
| FT | NON TER. |
| SQ | SEQUENCE 1 159585 MW; E3D9D08967724F07 CRC64; |
| | Query Match 28.6%; Score 1953.5; DB 5; Length 1433; |
| | Best Local Similarity 31.8%; Pred. No. 1.2e-138; |
| | Matches 465; Conservative 195; Mismatches 394; Indels 407; Gaps |
| QY | 26 CTGTDMKLRLPASPTLDMLRHLYQGQQVVGNNLELTYPNNSLSFLQDIOEVGYVL 85 |
| Db | 1 CIGTNGRMSPANREHYHKNLRDRYTCTNTYVDGNLEITWIGNITDLNFLOHIREVTGYVL 60 |
| QY | 86 IAHNVORVPQRLRIIVRGTLGF-----EDNYALAVLDNGDPLNNTTPVTGASPGGLREL 140 |
| Db | 61 ISLYDPQVILPRLQIIRGRITTKLNKEEAYGLFV-----SFSHMTL 104 |
| QY | 141 OLRSLTEILKGVLQIORNPOLCYOQTILWKDI-PHKNNOLATLIDNRSRACHPCSPMC 199 |
| Db | 105 ELPALRDILGSVGFNNYNLCHKMSINWEEILLAPQTSMQWTFFNSPERVCPCHPC 164 |
| QY | 200 KGSRCWGSESDQSCLTRTVCGAGCA--CKGPLPFNNFTVSFWLRVPKVSAASHL----- 252 |
| Db | 165 EVG-CWGEGAHNCRFSKLNCSPQCSQGRCFGPKP-----RECCHLCAGG 209 |
| QY | 253 ---EKHSCLACLPHNHSGICELHCPALVTYNTDTFESMPNEGTYTFGASCVTACPYN 309 |
| Db | 210 CTGPTQSDCLACKNFYDDGVCKQECPPMQIYNFTNYFWEPNDGKYAGATCVRKCP-EH 268 |
| QY | 310 LSTDVGSCTLCPLHNQVEATDGTRCEKSKPCARVCYGIGMEHLREVRAVTSANIQE 369 |
| Db | 269 LLKNGACVRKCPKGMQNSE-----CVPKGVCPKTCGGEGIVH-----SDNIGN 315 |
| QY | 370 FAGCKKIFGSLAFPESEDGPPASN-----APLOEQIQVFETLEEITGYLIYSAW 421 |
| QY | 316 YKDCIIIEGSLILDQSFDDGQQVYTNFSFGPRIKIIPDRILEVSTVKETIGFINIQAH 375 |

| | | | | |
|-----------------------|--------|---------------------|-------|--------------|
| Query Match | 28.6%; | Score 1953.5; | DB 5; | Length 1433; |
| Best Local Similarity | 31.8%; | Pred. No. 1.2e-138; | | |

Matches 465; Conservative 195; Mismatches 394; Indels 407; Gaps 38

26 CTGTDMLRLPASPETHLDMLRHLRYQGCQVQGNLELTLYLPTNASLSFLODIOEVQGYVL 85

4

I CIGINGRMSVPANREYHYKNLRDRYINCITYVDGNLEITWQNIITDLNLFQHIREVTGYVL 60

86 IAHNQVRQVPLQRLRIVRGTLF-----EDNYALAVLDNGDPLNNTTPTVGASPGGLREL 140

```
67 ISLYDLPOVILI.PRI.OIIRGRTEKLNKWEFA.YGIE.V-----SFESHMNTI. 104
```

THE UNIVERSITY OF CHICAGO

141 QLRSLTEILKGGVLIQRNPQLCYQDTILWKDI - FHKNNQLALTLDITNRSRACHPCSPMC 199

b 105 ELPALRDILGGSVGFNNYNLCHMK SINWEEILLAPQTS MQYTFNFSSPERVCP PCHPSC 164

200 KGSBCWGESSEDDCOSI.TBRTVCACCCA--PCKCPI.BENNETT/SEWI.PVDVISA.SHI.-----252

[illegible]

b 165 EVG-CWGEAHNCQRFSLNCSPPQCSQGRCFGPKP-----RECCHLFCAGG 209

253 ---EKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYN 309

210 CTCTPTGSDCTACINIEYDNDGVCYKQECDDMAIYADPTVSWPQNDPQCYAVYCTCTCTVYCR-EU 268

2 410 519F1Y9900CTACUAT1DD99VCNQ9ECFFMQTINFINIFMEFNFDFGNIAIGALCVKNCF-EH 208

310 LSTDVGSCTLVCP LHNQEVTAEDGTQRCCEKCSKPCARVCYGLGMEHLREVRAVTSANIQE 369

b
269 L L K D N G A C V R K C P K G K M P Q N S E - - - - - C V P C K G V C P K T C P G E G I V H - - - - - S D N I G N 315

370 BACCKKIECSI AEL DESEDCNDASNT - - - - - ADY ODEAY OVEETI EETTCVI VTSXW 421

076 TACOMMILCOCTRI DEESTDODFASNI - - - - - AFUFEQUVFIUEEIIIGIILISAM 421

: | | | | : | | | | : | | | | : | | | | :

: | | | | : | | | | : | | | | : | | | | :

b 316 YKDCTIIEGSLEILDQSFDFGQQVYTNFSFGPRYIKIDPDRLEVFSTVKEITGFINIAH 375

```
QY 422 POSLPDLSVFONLQVIRGRIHNGAY - SLTLOGLIGISWGLRLSLRLGSLGLALIHNTHL 480
Db 376 HPNFTLTNFRNLVVGROQLKENLASFVYIVKTSLSLEKLSLKRVSNGSIVILENSDL 435
QY 481 CFVHTVPDQOLFNRPHQALLHTANRDECEVCGGLACHQLCARGHCWGPGFTQCVCNSQF 540
Db 436 CFVEDIDWSEIKKSSDHEVMVQNRNATECHEEGMECSEQCSKAGCWGKGPEQCLECKNV 495
QY 541 LRQECVEECRVLQGLPREY - VNARHCLPCHPCPCOPNGSVTCFGPEADOCVACAHYKOP 599
Db 496 KYGKGLDSCK---SLPRLYSVDSKTCGDCHQCKD-----FCYGNEDNCGSCMNVKOG 547
QY 600 PFCVARCP-----SGVRPDLSSYP 618
Db 548 RFCVAECPTTHAMNGTCINCHKTCVCGRPRDTIAPDGCISCDKALIGSDAKIERCLMK 607
QY 619 INKFPD-----REG----- 627
Db 608 DESCPOGYVYDVYVLOEGLPLKQLSGAVCRKCHPRCKKCTGYGFHEQFCQECTGYKKGEQ 667
QY 628 -----ACQPCPINCT-----HSCVDL-----DD-----KG 647
Db 668 CEDECPODFYANEETRICLPCHQECRGCHGLGDDHHECRNLKLFEGDPYDNATFTCVSN 727
QY 648 CPAEQ-----RASPLTSIVSAVVGILLVVVLGVVFGI--- 679
Db 728 CPASHPYKRPQEQAGKIGPYCSADSMQSLRIEPTQVKIVMGSMVALILLCVVFGIAFV 787
QY 680 LTKRROOKIRKYMRLLOTELVELPTSGAMPNQAOIRLKETELRKVKVLGSGAFGT 739
Db 788 LFSRHKNKDAVAKNTWALACEDSEPLRPSNVGNPLTKRIIEASIRRGVGLGMGAFGR 847
QY 740 VYGIWIPGGENVKIPVAIKVLRENTSPKANKIILDEAYVMAGVGPYVSRLLGICLTST 799
Db 848 VFKGVMPGESVKIPVAIKVLMMSGSSSEKFLAEAYIMASVEHPNLLKLLAVCMTSQ 907
QY 800 VOLVQLMPYGCILLDHVRENRLGSGODLLNWCQIAKMSYLEDVRLVHRDLAARNVLV 859
Db 908 MMLITQLMPLGLCLDYVRNKKIGSKALLNWSQTARGMAYLEERLVRHDLAARNVLV 967
QY 860 KSPNHVKITDFGLARLLIDETEHYHADGKGVPIKWALESILRRRTHQSDVWSYGVTVW 919
Db 968 QTPSCVKITVFLGAKLLDFDSDEYRAAGGKMPKWLALCEIRHRVFTSKSDVMWAFGITW 1027
QY 920 ELMTFCAKPYDGIPIAREIPDLLEKGERLPPOPCTIDVYMWKMWIDSECRPRFELV 979
Db 1028 ELLTYGARPYENVYPAKDVPELIEIGHKLPOPDICSLDVYILLSCWVLDADARPTFKLA 1087
QY 980 SEFSRMARDPQRFVWIONEDLGPASPLDSTFYRSLLEDDDMGDLV----- 1024
Db 1088 ETPAEKARDPGRYLM-----PDGKFMRLPSYTNQDEKDLIRTLAPVMAAAAAA 1138
QY 1025 -----DAEEYLVPOQGFPCPDPAAGGWVHRRSSSTRSGGDLTLGLE 1070
Db 1139 AGASNVDPSTIAETDEYLOPKTRPSIMLPGSA-----VE 1175
QY 1071 PSEEEAPRS-----PLAP---SEGAGSDVFDGLGMGAAGKQLSLPHTHPSPLQRYSED 1121
Db 1176 PS-DEMPKSLRYKCDPLKPDDETGDHKEV-----GVGGIR----- 1210
QY 1122 PTVPLSETDGYVAPLTCSPQRYVNOQDVORPOPPSPREGPLPAARPAGATLERAKTLSP 1181
Db 1211 --LNLPLDEDDVLLMP-TCOSQ---NQS----- 1233
QY 1182 GRKNGVKDVFAFGAVENBEYL-----TPQGGNAPQHPHPPAFSPAFDNLVYWDQD 1232
Db 1234 G-----YMDLIGVPASVDNBEYLMGMSQTQIAGLAQSGMG---PHTPP----- 1272
QY 1233 PPERGAPSTFKGTPTAENPE 1253
Db 1273 -----PPNTPNGMPTHQHSQ 1287
```

RESULT 8

Q9UK79

ID Q9UK79 PRELIMINARY; PRT; 419 AA.

AC Q9UK79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Heretatin.

GN HER-2.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]
SEQUENCE FROM N.A.

RX MEDLINE=99415951; PubMed=10485918;

RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;

RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted

RT autoinhibitor.";

RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).

RN [2]
SEQUENCE FROM N.A.

RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF177761; AAD56009.2; -

DR InterPro: IPR000494; EGFR_L domain.

DR InterPro: IPR002174; Furin-like.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF01030; Recep_L domain; 1.

DR SMART; SM00261; FU; 1.

SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match

Best Local Similarity

Matches 339; Conservative 6; Mismatches 34; Indels 29; Gaps 4;

QY 1 MELAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPEHLDMLRHLHYQGCVVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPEHLDMLRHLHYQGCVVQGNL 60

QY 61 ELYLPNTASLFLQDIEVQGVLIHNVQVPLORLIRVGTQLFEDNYALAVLDNG 120

Db 61 ELYLPNTASLFLQDIEVQGVLIHNVQVPLORLIRVGTQLFEDNYALAVLDNG 120

QY 121 DPLNNTPTVGTASPGGLRELQRLSLEILKGGVLIQORNPOLCYQDTILWKDIFHKNOLA 180

Db 121 DPLNNTPTVGTASPGGLRELQRLSLEILKGGVLIQORNPOLCYQDTILWKDIFHKNOLA 180

QY 181 LTLIDTNRGRACHPCSPMCKGSRGCGESSEDCQSLTRTVCCAGGCARCKGPLPFTVSF 240

Db 181 LTLIDTNRGRACHPCSPMCKGSRGCGESSEDCQSLTRTVCCAGGCARCKGPLPFTVSF 240

QY 241 WLRVPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 292

Db 233 -----TDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 286

QY 293 GRVTFGASCVTACPNYLVSTDVGSCVTLVCPHMQEVTAEQGTORCKCKKPCARVCYGLG 352

Db 287 GRVTFGASCVTACPNYLVSTDVGSCVTLVCPHMQEVTAEQGTORCKCKKPCARVCYGLG 341

QY 353 MEHLREVAVTSANIOEFAGCKKIFGSLAPLPESFDGDPASNTAPLQ 400

Db 342 THSLPRPRAVPVPLRMQPG--PAHPVLSFLRPSWDLVSFAFYSLPLAP 387

RESULT 9

Q8R2X1

ID Q8R2X1 PRELIMINARY; PRT; 367 AA.

AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Hypothetical 40.2 kDa protein.

```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027080; AAH27080.1; -.
KW Hypothetical protein.
SQ
  895 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPPICT 954
  1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPPICT 60
  955 IDVYIMVWCWIDSECRPRFRELVSFPRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1014
  61 IDVYIMVWCWIDSECRPRFRELVSFPRMARDPQRFVVIQNEIDLGPSSPMDSTFYRSL 120
  1015 LEDDDMGDLVDAEYLVPOGGFCPPAPGACGMVHRRSSSTRSGGDLTLGLEPSEE 1074
  121 LEDDDMGDLVDAEYLVPOGGFCPPALGTGSTAHRHRRSSARSGGELTLGLEPSEE 180
  1075 EAPRSPAPSEAGSDVFDGDLGMAAGLQSLTHDPSPLOQRYSDPTPLPSETDGYV 1134
  181 EAPRSPAPSEAGSDVFDGLAVGYTKGLQSLSPHDLSPLOQRYSDPTPLPSETDGYV 240
  1135 APLTCSPOEYVQNPQVPRQPPSPREGPLPAARPAGATLERAKTLPSPKNGVVKVFAFG 1194
  241 APLTCSPOEYVQNPQVPRQPPSPREGPLPAARPAGATLERAKTLPSPKNGVVKVFAFG 300
  1195 GAVENPEYLTPOGGAAPQHPAPSPAFDNLVYWDQPPERCAPSTFKGTPTAENPEY 1254
  301 GAVENPEYLTPOGGAAPQHPAPSPAFDNLVYWDQPPERCAPSTFKGTPTAENPEY 360
  1255 LGLDVPV 1261
  361 LGLDVPV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
  Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
  evolution of distinct viral genomes carrying mutant v-erbB genes with
  different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSSP; P03322; 1A6S
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.
```

```
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ
  575 PQNGSVTCFPGPADQCVACAHYKDPFCVACRPSGVKPDLSYMPIKFPDEEGACQCPPI 634
  141 PEETATPTKGP--DHCCKAHFDGPHCVKACPAVLGENDTL-VMKYADANAVCQLCHP 197
  635 NCTHSCVDLDDKGCAPAEQASPLTSTVSAVV--GILLVVLGVVGVFGILIKRQOKIRKVTM 693
  198 NCTRGCKGGLGEGCP---NGSKTPSTAAGVVGGLLCLLVVGLGIGLYLRRR-HIVRKRTL 253
  694 RRLQETELVEPLTPSGAMPNOAQRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVK 753
  254 RRLQERELVEPLTPSGEAPNOAHLRLAKETEFKVKVLGSGAGFTVYKGLWIPGEKVK 313
  754 IPVAIKVLRENTSPKANKEILDEAYVMAGVSPYSRLGICLTSTVQLVTQMPYGCLL 813
  314 IPVAIKELREATSPKANKEILDEAYVMASVNDPRVCRLLGICLTSTVQLITQMPYGCLL 373
  814 DHVRENRLGSGDQLLNCWQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 873
  374 DYIREHKONIGSQYLLNMCVQIAKGNVLEERLVRDLAARNVLVKTPOHVKITDFGLA 433
  874 RLDDIDETEHADGGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIP 933
  434 KLGDAKEYHAEGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIP 493
  934 AREIPDLLEKGERLPQPPICTIDVYIMVWCWIDSECRPRFRELVSFPRMARDPQRFV 993
  494 ASEISVLEKGERLPQPPICTIDVYIMVWCWIDADSRPKFRELIAEFSKMARDPPEYL 553
  994 VIQ-NEDLGPASPLDSTFYRSLDDDDMGDLVDAEYLVPOGGFCPPAPGACGMVHRR 1052
  554 VIQGERMHLPTDTSKFTYRLMEEDMEDIVDADSYLVPHQGGF----- 598
  1053 HRSSSTRSGGDLTLGLEPSEEBEAPRSP-----APSEAGSDVFDGDLGMAAGLQSL 1107
  599 -NSPST-----SRTPLLSLSATSNNSATNCID-----RNCQGH 631
  1108 PTHDPSPLORYSEDPTVPLPSET--DGYVAPLTCSPQYVNPQDVRPQPPSPREGPLPA 1165
  632 PVREDSFVQRYSSDPTGNFLESSIDDGFL-----PAPEYVQ--LMPKKPS----- 675
  1166 ARPAGATLERAKTLPSPKNGVVKVDFP-----APFGAVENPEYL 1203
  676 -----TAMVQNIYNNISLTAISKLPMDSRVQNSHSTAVDNPPEYL 715

RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
```

RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS0011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 25.2%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 2e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 584 GPEADQCACAHYKDPFCVARGSGYKPDLSYMPIWKFPDEBAGACQPCPINTHSCVDL 643
Db 1 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCQLCHPNCTRGCKGP 57

QY 644 DDKGCCAEQASPLTSIVSAV-GILLVVLGVVFGILLIKERQOKIRKYMRLLQETEL 702
Db 58 GLEGGP---NGSKTPSIAAGVGGGLVCLVVGIGLYLRRR-HIVKRTLRLLQRELL 113

QY 703 VEPLTPSGAMPNOAQRILKETELRKVKVLGSGAGFYVYKGIWIPDGENVKIPVAIKVLR 762
Db 114 VEPLTPSGEAPNOAHLILKETEFKVKVLGSGAGFYVYKGLWIPGEKVKIPVAIKELR 173

QY 763 ENTSPKANKEILDEAYVMAGVGSYVSRLLGICLTSTVQLVTLQMPYGCILLDHVRENRR 822
Db 174 EATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLVTLQMPYGCILLDVIREHKN 233

QY 823 LGSODLLNWCQIAKNSYLEDLVHRDLAARNVLKSPNHVKITDFGLARLLDDETE 882
Db 234 IGSOYLLNWCQIAKSNYLEERLVRDLAARNVLKTPQHKVITDFGLAKLLGADKE 293

QY 883 YHADGGKVPKMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLLE 942
Db 294 YHAEGKVPKMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353

QY 943 KGERLPQPPCTIDVYIMVKWMDSECRPRFRELVSFERSMRARDPQRFVVIQ-NEDLG 1001
Db 354 KGERLPQPPCTIDVYIMVKWMDADSRPKFRELIAEFSKMARDPYRLVIOGDERMH 413

QY 1002 PASPLDSTFYRSLLEDGDLVDAEYLVPOQGFCCPDPAAGAGVHHRHSSSTRSG 1061
Db 414 LPSPTUSKFRTLMEEDMEDIVDADRYLVPHOGFF-----NSPST--- 454

QY 1062 GGDLTGLLEPSEEAAPRSL-----APSEAGSDVDFDGLGMAAKGLQLSLPHDPSPLQ 1116
Db 455 -----SRTPLLSLSLATSNNATNCID-----RNGQGHVREDVSFVQ 491

QY 1117 RYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNPQDVPRPQPPSPREGPLPAARAGATLE 1174
Db 492 RYSDPTGNFLEESIDGFL-----PAPEYVNO--LMPKKPS----- 526

QY 1175 RAKTLSPGKGVKVDVF-----AFGGAVENPEYL 1203
Db 527 ----TAMVQNIYNNISLTAKLPMDSRYQNSHSTAVDNPEYL 566

RESULT 12
Q8WYV0
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN PP3659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00108; EF HAND; UNKNOWN 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 4.5e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 895 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLLEKGERLPQPPICT 954
Db 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLLEKGERLPQPPICT 60

QY 955 IDVYIMVKWMDISECRPRFRELVSFERSMRARDPQRFVVIQEDLGPASPLDSTFYRSL 1014
Db 61 IDVYIMVKWMDISECRPRFRELVSFERSMRARDPQRFVVIQEDLGPASPLDSTFYRSL 120

QY 1015 LEDDDMGDLVDAEYLVPOQGFCCPDPAAGAGVHHRHSSSTRSGGDLTLGLEPSEE 1074
Db 121 LEDDDMGDLVDAEYLVPOQGFCCPDPAAGAGVHHRHSSSTRSGGDLTLGLEPSEE 180

QY 1075 EAPRSLAPSEAGSDVDFDGLGMAAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGYV 1134
Db 181 EAPRSLAPSEAGSDVDFDGLGMAAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGYV 240

QY 1135 APLTCSPOPEYVNPQDVPRPQPPSPREGPLPAARAGATLERAKTLSPGKGVKVDVFAFG 1194
Db 241 APLTCSPOPEYVNPQDVPRPQPPSPREGPLPAARAGATLERAKTLSPGKGVKVDVFAFG 300

QY 1195 GAVENPEYLTPOGGAAPQP-----HPPPA---FSPAFDNL 1226
Db 301 GAVENPEYLTPOGGAALSPTLLPSAQPSSTTSITGTRTHQSGGLHPAPSKGHLRQRTQST 360

QY 1227 YYWD-QDPPER-----GAPPTFKGTPTAEN 1251
Db 361 WWVTCQEPGEQVRRSPDVSSSGSREGLTSAGIKRWEGPPTTSRGTCARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN GAG, v-erb-A, v-erb-B.
OS Avian erythroblastosis virus.

FT Region /label= Cysteine_rich_domain
210..224
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
250..264
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Domain
324..483
/label= Ligand_binding_domain
FT Region
325..339
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
369..383
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
465..479
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Domain
484..623
/label= Cysteine_rich_domain
FT Region
579..593
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Domain
624..654
/label= Transmembrane_domain
FT Region
632..652
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
653..667
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Domain
655..1010
/label= Tyrosine_kinase_domain
FT Region
661..675
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
695..709
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
710..730
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Domain
1011..1235
/label= C-terminal_domain
XX
WO200020027-A2.
XX
13-APR-2000.
XX
05-OCT-1999; 99WO-DK00525.
XX
05-OCT-1998; 98DK-0001261.
XX
20-OCT-1998; 98US-0105011.
XX
(MEBI-) M & E BIOTECH AS.
XX
Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
PI Gautam A, Birk P, Karlsson G;
XX
WPI; 2000-349917/30.
XX
N-PSDB; AAA09455.
XX
Inducing immune responses to weakly immunogenic, tumor associated
PT peptide antigens for the treatment of breast and prostate cancer
XX
XX
Claim 62; Page 193-198; 220pp; English.
XX
This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
CC response. Subdominant CTL epitopes, antibody binding regions and
CC cysteine residues involved in disulfide bonds are preserved in the
CC immunogenized forms. Regions suitable for the insertion of foreign T
CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
CC cell-associated peptide antigens (PA) such as those associated with
CC cancers (self-proteins), e.g. human prostate specific membrane antigen
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC The method comprises effecting simultaneous presentation by antigen
CC producing cells (APCs) of the animals immune system of: (1) at least 1
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC B-cell group derived from the cell-associated PA; and (2) at least 1
CC first T helper cell group which is foreign to the animal. Analogues of
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC part of all known and predicted CTL and B-cell epitopes of the respective
CC PA and including at least one foreign T helper epitope are also claimed.
CC The method is used to treat prostate, prostate/breast or breast cancer
CC when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ Sequence 1255 AA;

Query Match 98.0%; Score 6690; DB 21; Length 1255;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1241; Conservative 1; Mismatches 5; Indels 22; Gaps 2;
Qy 1 MELAALCRWGLLLALLPFGAASCTGCTDMKRLPASPETHLDMRLHYOCQVVGNL 60
Db 1 MELAALCRWGLLLALLPFGAASCTGCTDMKRLPASPETHLDMRLHYOCQVVGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Qy 121 DPLNNTPTVTGASPGGLREQLRSITELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTPTVTGASPGGLREQLRSITELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLDTNRSRACHPCSPMKGSRGWESSEDCOSITRTVCAGGCARCKGKPLFNFTVSF 240
Db 181 LTLDTNRSRACHPCSPMKGSRGWESSEDCOSITRTVCAGGCARCKGKPLFNFTVSF 240
Qy 241 WLRVPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 286
Qy 293 GRYTFGASCVTACPNYLSLTDVGSCTLVCPLNHQVTAEDGTQRCCKSPCARCYGIG 352
Db 287 GRYTFGASCVTACPNYLSLTDVGSCTLVCPLNHQVTAEDGTQRCCKSPCARCYGIG 346
Qy 353 MEHLREVRVAVTSANIOEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQLOVFTLEEI 412
Db 347 MEHLREVRVAVTSANIOEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQLOVFTLEEI 406
Qy 413 TGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRSLRELGSGLA 472
Db 407 TGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRSLRELGSGLA 466
Qy 473 LIHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPT 532
Db 467 LIHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPT 526
Qy 533 QCVNCSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 592
Db 527 QCVNCSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 586
Qy 593 CAHYKDPFPFCVARCPGKVPDLSYMPWKFPDEGACQPCPINCCTHSCVDLDDKGCBAEQ 652
Db 587 CAHYKDPFPFCVARCPGKVPDLSYMPWKFPDEGACQPCPINCCTHSCVDLDDKGCBAEQ 646
Qy 653 RASPLTSIVSAVGIILLVVLGVVFGIILIKRQOKIRKYTWRRLLQETLEVEPLTPSGAM 712
Db 647 RASPLTSIVSAVGIILLVVLGVVFGIILIKRQOKIRKYTWRRLLQETLEVEPLTPSGAM 706
Qy 713 PNOQMRLKETELRKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKE 772
Db 707 PNOQMRLKETELRKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKE 766

```
QY 773 ILDEAYVMAGVGSPPVSRLLIGICLTSTVQLVTOLMPYGCCLLDHVRNRRGLGSODLLNWC 832
Db 767 ILDEAYVMAGVGSPPVSRLLIGICLTSTVQLVTOLMPYGCCLLDHVRNRRGLGSODLLNWC 826
QY 833 MQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 892
Db 827 MQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 886
QY 893 KMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLP 952
Db 887 KMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLP 946
QY 953 CTIDVYMWKMWIDSECRPRFRELVSERFMRDPOQFVVIQNEIDLGPASPLDSTFYR 1012
Db 947 CTIDVYMWKMWIDSECRPRFRELVSERFMRDPOQFVVIQNEIDLGPASPLDSTFYR 1006
QY 1013 SLEDDDDMGDLVDAEYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPS 1072
Db 1007 SLEDDDDMGDLVDAEYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPS 1066
QY 1073 EEEAPRSLAPSEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVP 1132
Db 1067 EEEAPRSLAPSEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVP 1126
QY 1133 YVAPLTCSPOEYVYVQDPVRPOPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVA 1192
Db 1127 YVAPLTCSPOEYVYVQDPVRPOPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVA 1186
QY 1193 FGGAIVENPEYLTPOGGAAPQHPHPPAFSPADNLYYWDQDPPERGAPPSTFKGTPTAENP 1252
Db 1187 FGGAIVENPEYLTPOGGAAPQHPHPPAFSPADNLYYWDQDPPERGAPPSTFKGTPTAENP 1246
QY 1253 EYGLGDVPV 1261
Db 1247 EYGLGDVPV 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
AC AAE12130;
DT 18-DEC-2001 (first entry)
XX Human tyrosine kinase-type receptor, HER-2.
DE
XX Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX Region 774..782
XX /note= "Antigenic epitope"
XX
XX WO200168677-A2.
XX
XX 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX 16-MAR-2000; 2000US-0527487.
XX
XX (GENZ ) GENZYME CORP.
XX
XX Nicolette CA;
XX
XX WPI; 2001-616284/71.
XX
XX N-PSDB; AAD19731.
XX
```

```
PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT
XX Claim 4; Page 63-67; 69pp; English.
XX
XX The invention relates to synthetic therapeutic compounds (antigenic
XX peptides) with enhanced binding to major histocompatibility complex
XX (MHC) molecules and enhanced immunoregulatory properties relative
XX to their natural counterparts. Compounds of the invention are useful
XX for inducing an immune response in a subject and for use in adoptive
XX immunotherapy. They are useful as components of anti-cancer vaccines
XX and to expand immune effector cells that are specific for cancers
XX characterised by expression of the breast cancer antigen, HER-2.
XX Polynucleotides that encode peptides of the invention are useful as
XX hybridisation probes and as primers for the detection of genes of gene
XX transcripts that are expressed in antigen presenting cells (APCs), to
XX confirm transduction of polynucleotides into host cells. The present
XX sequence is human tyrosine kinase-type receptor, HER-2. Compounds
XX of the invention are designed based on the HER-2 antigenic peptide
XX (774-782).
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6690; DB 22; Length 1255;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1241; Conservative 1; Mismatches 5; Indels 22; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASTOVCTGDMKRLPASPETHLDMLRHLGYCCVVOGNL 60
Db 1 MELAALCRWGLLALLPPGAASTOVCTGDMKRLPASPETHLDMLRHLGYCCVVOGNL 60
QY 61 ELTYLPTNASLSFLQIQEVQYVLIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQIQEVQYVLIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVGAGGACRCKGPIPFNNFTVSF 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVGAGGACRCKGPIPFNNFTVSF 232
QY 241 WLRVPKVSASHLE-----KHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPE 286
QY 293 GRVTFGASCVTACPYNYLSTDVSGCTLVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLG 352
Db 287 GRVTFGASCVTACPYNYLSTDVSGCTLVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLG 346
QY 353 MEHLREVRATVSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFTLBEI 412
Db 347 MEHLREVRATVSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFTLBEI 406
QY 413 TGYLYISAWPDSLPLDSVFQNLQVIRGRIILHNGAYSILTLQGLGISWLGRLSRLRELGSLA 472
Db 407 TGYLYISAWPDSLPLDSVFQNLQVIRGRIILHNGAYSILTLQGLGISWLGRLSRLRELGSLA 466
QY 473 LIHNHNLFCFVHTVPDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHGWGPGPT 532
Db 467 LIHNHNLFCFVHTVPDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHGWGPGPT 526
QY 533 QCVNCSQFLRGQECVBEICRVLQGLPRBYVNAHCLPCHPECPQNGSVTCFPGBADQCV 592
Db 527 QCVNCSQFLRGQECVBEICRVLQGLPRBYVNAHCLPCHPECPQNGSVTCFPGBADQCV 586
QY 593 CAHYKDPFPCVARCPSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCRAEQ 652
Db 587 CAHYKDPFPCVARCPSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCRAEQ 646
```

```
PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT
XX Claim 4; Page 63-67; 69pp; English.
XX
XX The invention relates to synthetic therapeutic compounds (antigenic
XX peptides) with enhanced binding to major histocompatibility complex
XX (MHC) molecules and enhanced immunoregulatory properties relative
XX to their natural counterparts. Compounds of the invention are useful
XX for inducing an immune response in a subject and for use in adoptive
XX immunotherapy. They are useful as components of anti-cancer vaccines
XX and to expand immune effector cells that are specific for cancers
XX characterised by expression of the breast cancer antigen, HER-2.
XX Polynucleotides that encode peptides of the invention are useful as
XX hybridisation probes and as primers for the detection of genes of gene
XX transcripts that are expressed in antigen presenting cells (APCs), to
XX confirm transduction of polynucleotides into host cells. The present
XX sequence is human tyrosine kinase-type receptor, HER-2. Compounds
XX of the invention are designed based on the HER-2 antigenic peptide
XX (774-782).
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6690; DB 22; Length 1255;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1241; Conservative 1; Mismatches 5; Indels 22; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASTOVCTGDMKRLPASPETHLDMLRHLGYCCVVOGNL 60
Db 1 MELAALCRWGLLALLPPGAASTOVCTGDMKRLPASPETHLDMLRHLGYCCVVOGNL 60
QY 61 ELTYLPTNASLSFLQIQEVQYVLIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQIQEVQYVLIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVGAGGACRCKGPIPFNNFTVSF 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVGAGGACRCKGPIPFNNFTVSF 232
QY 241 WLRVPKVSASHLE-----KHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPE 286
QY 293 GRVTFGASCVTACPYNYLSTDVSGCTLVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLG 352
Db 287 GRVTFGASCVTACPYNYLSTDVSGCTLVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLG 346
QY 353 MEHLREVRATVSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFTLBEI 412
Db 347 MEHLREVRATVSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFTLBEI 406
QY 413 TGYLYISAWPDSLPLDSVFQNLQVIRGRIILHNGAYSILTLQGLGISWLGRLSRLRELGSLA 472
Db 407 TGYLYISAWPDSLPLDSVFQNLQVIRGRIILHNGAYSILTLQGLGISWLGRLSRLRELGSLA 466
QY 473 LIHNHNLFCFVHTVPDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHGWGPGPT 532
Db 467 LIHNHNLFCFVHTVPDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHGWGPGPT 526
QY 533 QCVNCSQFLRGQECVBEICRVLQGLPRBYVNAHCLPCHPECPQNGSVTCFPGBADQCV 592
Db 527 QCVNCSQFLRGQECVBEICRVLQGLPRBYVNAHCLPCHPECPQNGSVTCFPGBADQCV 586
QY 593 CAHYKDPFPCVARCPSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCRAEQ 652
Db 587 CAHYKDPFPCVARCPSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCRAEQ 646
```

Qy 653 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 712
Db |||||||
Qy 647 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 706
Db |||||||
Qy 713 PNOAQMRLKETELRKVKVILGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 772
Db |||||||
Qy 707 PNOAQMRLKETELRKVKVILGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 766
Db |||||||
Qy 773 ILDEAYVMAGVSPVSRLLGICITSTVQLVTLQMPYGCILLDHDVRENRRGLSQDILLNWC 832
Db |||||||
Qy 767 ILDEAYVMAGVSPVSRLLGICITSTVQLVTLQMPYGCILLDHDVRENRRGLSQDILLNWC 826
Db |||||||
Qy 833 MQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 892
Db |||||||
Qy 827 MQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 886
Db |||||||
Qy 893 KWMALLESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPI 952
Db |||||||
Qy 887 KWMALLESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPI 946
Db |||||||
Qy 953 CTIDVYIMVKCWMIDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYR 1012
Db |||||||
Qy 947 CTIDVYIMVKCWMIDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYR 1006
Db |||||||
Qy 1013 SLEDDDDMGDLVDAEYLVPOQGFPCPDPAAGGMVHRRSSSTRSGGDLTLGLEPS 1072
Db |||||||
Qy 1007 SLEDDDDMGDLVDAEYLVPOQGFPCPDPAAGGMVHRRSSSTRSGGDLTLGLEPS 1066
Db |||||||
Qy 1073 EEEAPRSLAPSEGAGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDG 1132
Db |||||||
Qy 1067 EEEAPRSLAPSEGAGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDG 1126
Db |||||||
Qy 1133 YVAPLTCSPQPEYVYNQDVRPQPPSPREGPLPAARPAGATILERAKTLSPGKNGVVKDVA 1192
Db |||||||
Qy 1127 YVAPLTCSPQPEYVYNQDVRPQPPSPREGPLPAARPAGATILERAKTLSPGKNGVVKDVA 1186
Db |||||||
Qy 1193 FGGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENP 1252
Db |||||||
Qy 1187 FGGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENP 1246
Db |||||||
Qy 1253 EYGLGDVPV 1261 /
Db |||||||
Qy 1247 EYGLGDVPV 1255
Db |||||||
RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
XX AAB60167;
DT 03-APR-2001 (first entry)
XX HER2 transgene plasmid construct encoded protein.
DE Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
KW Homo sapiens.
XX Synthetic.
OS WO200100244-A2.
XX PN 04-JAN-2001.
XX PD 23-JUN-2000; 2000WO-US17229.
XX PF 25-JUN-1999; 99US-0141316.
XX PR 16-MAR-2000; 2000US-0189844.
XX PA (GETH) GENENTECH INC.
XX Erickson S, Schwalli R;
PI

XX WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6690; DB 22; Length 1255;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1241; Conservative 1; Mismatches 5; Indels 22; Gaps 2;
Qy 1 MELAALCRWGLLLALLPPGAASCTVCTGDMKRLPASPETHDMLRHLHYGQCQVVGNL 60
Db |||||||
Qy 1 MELAALCRWGLLLALLPPGAASCTVCTGDMKRLPASPETHDMLRHLHYGQCQVVGNL 60
Db |||||||
Qy 61 ELTYLPTNASLSFLQDIOEVQYVLIHNOVROPLORLRVRGTLQFEDNYALAVLNG 120
Db |||||||
Qy 61 ELTYLPTNASLSFLQDIOEVQYVLIHNOVROPLORLRVRGTLQFEDNYALAVLNG 120
Db |||||||
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILWKDIFHKNQLA 180
Db |||||||
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGVLIQRNPOLCYODTILWKDIFHKNQLA 180
Db |||||||
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWSSESQSLTRTVACAGGCARCKGPIPFNNFTVSF 240
Db |||||||
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWSSESQSLTRTVACAGGCARCKGPIPFNNFTVSF 232
Db |||||||
Qy 241 WLRVPKVSASHLE-----KHSDDLACLFHNSGICELHCPALVTYNTDTFESMPNPE 292
Db |||||||
Qy 233 -----TDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 286
Db |||||||
Qy 293 GRYTFGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLG 352
Db |||||||
Qy 287 GRYTFGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLG 346
Db |||||||
Qy 353 MEHLREVRVTSANIQBFAGCKITFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEBI 412
Db |||||||
Qy 347 MEHLREVRVTSANIQBFAGCKITFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEBI 406
Db |||||||
Qy 413 TGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELSGLA 472
Db |||||||
Qy 407 TGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELSGLA 466
Db |||||||
Qy 473 LIHNTLHLCFVHTVPWDLFRNPHQALLHTANRPEDECVGEGLAGCHQOLCARGHCHGPGPT 532
Db |||||||
Qy 467 LIHNTLHLCFVHTVPWDLFRNPHQALLHTANRPEDECVGEGLAGCHQOLCARGHCHGPGPT 526
Db |||||||
Qy 533 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 592
Db |||||||
Qy 527 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 586
Db |||||||
Qy 593 CAHYKDPFPFCVARCPGVKPDLSYMPWKFPDESGACQPCINCTHSCVDLDDKGCPEAQ 652
Db |||||||
Qy 587 CAHYKDPFPFCVARCPGVKPDLSYMPWKFPDESGACQPCINCTHSCVDLDDKGCPEAQ 646
Db |||||||
Qy 653 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 712
Db |||||||
Qy 647 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 706
Db |||||||
Qy 713 PNOAQMRLKETELRKVKVILGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 772
Db |||||||

Db 707 PNQAMRILKETELKVKVILGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 766
Qy 773 ILDEAYVMAGVSPVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGRLGSDQLLNCW 832
Db 767 ILDEAYVMAGVSPVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGRLGSDQLLNCW 826
Qy 833 MQIAKGSYLEDVRLVHRDLAARNVVKSPHVKITDFGLARLLDIDITEYHADGKVP 892
Db 827 MQIAKGSYLEDVRLVHRDLAARNVVKSPHVKITDFGLARLLDIDITEYHADGKVP 886
Qy 893 KMALESILRRRFTHOSDVMWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 952
Db 887 KMALESILRRRFTHOSDVMWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 946
Qy 953 CTIDVYMIWKCWIMIDSECRPRFRELVSFSEFARMARDPQRFVVYIQNEDLGSPALDSTFYR 1012
Db 947 CTIDVYMIWKCWIMIDSECRPRFRELVSFSEFARMARDPQRFVVYIQNEDLGSPALDSTFYR 1006
Qy 1013 SLEDDDMGDLVDAEYLVPOQGFPCPDPAAGAGWVHRRSSSTRSGGDLTGLBPS 1072
Db 1007 SLEDDDMGDLVDAEYLVPOQGFPCPDPAAGAGWVHRRSSSTRSGGDLTGLBPS 1066
Qy 1073 EEEAPRSLAPSEAGSDVDFDGLCMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDG 1132
Db 1067 EEEAPRSLAPSEAGSDVDFDGLCMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDG 1126
Qy 1133 YVAPLTCSPQBYNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVDVFA 1192
Db 1127 YVAPLTCSPQBYNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVDVFA 1186
Qy 1193 FGGAVENPEYLTPOGGAAPQHPHPPAFSPADNLYYWDODDPERGAPSTFKGTPTAENP 1252
Db 1187 FGGAVENPEYLTPOGGAAPQHPHPPAFSPADNLYYWDODDPERGAPSTFKGTPTAENP 1246
Qy 1253 EYLGLDVVPV 1261
Db 1247 EYLGLDVVPV 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
AC AAU74545;
XX
XX
XX 23-APR-2002 (first entry)
XX Human HER2 (Erbb2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macropahagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
XX US2002001587-A1.
XX
XX 03-JAN-2002.
XX
XX 16-MAR-2001; 2001US-0811123.
XX
XX 16-MAR-2000; 2000US-189844P.
XX
XX 05-OCT-2000; 2000US-238327P.
XX
XX (ERIC/) ERICKSON S.
XX (SCHW/) SCHWALL R.
XX (SLIW/) SLIWOWSKI M.
XX

PI Erickson S, Schwall R, Sliwowski M;
XX WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
is characterised by the overexpression of an epidermal growth factor
receptor (Erbb) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
neural, glial, astrocytal, hypothalamic, glandular, macropahagal,
CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
immunological disorders. This sequence represents the human HER2 (Erbb2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6690; DB 23; Length 1255;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1241; Conservative 1; Mismatches 5; Indels 22; Gaps 2;
Qy 1 MELAALCRWGLLLALLPPGAASSTVCTGDMKRLPASPETHLMLRLHYQCCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTVCTGDMKRLPASPETHLMLRLHYQCCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEQVQYVLIHQNVRQVPLQRLRVRGTLFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIEQVQYVLIHQNVRQVPLQRLRVRGTLFEDNYALAVLNG 120
Qy 121 DPLNNTPTVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTPTVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGRCWGSSEDCQSLTRTVCAAGCARCGPLPFFNFTVSF 240
Db 181 LTLIDTNRSRACHPCSPMKGRCWGSSEDCQSLTRTVCAAGCARCGPLP----- 232
Qy 241 WLRVPKVSASHLE-----KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPE 286
Qy 293 GRVTFGASCVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLG 352
Db 287 GRVTFGASCVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLG 346
Qy 353 MEHLREVRVTSANIQEFAGCKIFGSLAPFESFDGDPASNTAPLOEQLOVFTLEEI 412
Db 347 MEHLREVRVTSANIQEFAGCKIFGSLAPFESFDGDPASNTAPLOEQLOVFTLEEI 406
Qy 413 TGYLYISAWPDSLPLDSVFQNLQVIRGRIHLHNGAYSLTQGLGISWGLRSLRELGSGLA 472
Db 407 TGYLYISAWPDSLPLDSVFQNLQVIRGRIHLHNGAYSLTQGLGISWGLRSLRELGSGLA 466
Qy 473 LIHNTHLCFVHTVPWDQLFRNPHQALLHTANRDECEVGEGLACHQICARGHCWGPGPT 532
Db 467 LIHNTHLCFVHTVPWDQLFRNPHQALLHTANRDECEVGEGLACHQICARGHCWGPGPT 526
Qy 533 QCVNCSOFLRGQECVBEICRVLQGLPREVVARHCLPCHPECPONGSVTCFGEADQVCA 592
Db 527 QCVNCSOFLRGQECVBEICRVLQGLPREVVARHCLPCHPECPONGSVTCFGEADQVCA 586
Qy 593 CAHYKDPFPFCVARPCSGVKPDLSPYMPWKFPDEBGAQCPQPCINCTHSCVDLDDKGCPEQ 652


```
Db 587 CAHYKDPFCVAPCSGKVPDLSPYPIWKFPDEGACQPCPINCTSHSCVDLDDKCPAEQ 646
QY 653 RASPLTSIVSAVVGILLVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAM 712
Db 647 RASPLTSIVSAVVGILLVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAM 706
QY 713 PNOQMIRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNRENTSPKANKE 772
Db 707 PNOQMIRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNRENTSPKANKE 766
QY 773 ILDEAYVMAGVSGPVVSELLGICLTSTVOLVTLQMPYGCCLLDHVRNRRGRLGSDQLLNCW 832
Db 767 ILDEAYVMAGVSGPVVSELLGICLTSTVOLVTLQMPYGCCLLDHVRNRRGRLGSDQLLNCW 826
QY 833 MOIAGMSYLEDVRLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPPI 892
Db 827 MOIAGMSYLEDVRLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPPI 886
QY 893 KMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPPI 952
Db 887 KMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPPI 946
QY 953 CTIDVYMIWVKWIMIDSCRRFRRELVSFSEFRMARDPQRFVVIQNEIDLGPASPLDSTFYR 1012
Db 947 CTIDVYMIWVKWIMIDSCRRFRRELVSFSEFRMARDPQRFVVIQNEIDLGPASPLDSTFYR 1006
QY 1013 SLEDDDDMGDLVDABEYLVPQGGFCPPAPGAGGMVHRRSSSTRSGGDLTLGLEPS 1072
Db 1007 SLEDDDDMGDLVDABEYLVPQGGFCPPAPGAGGMVHRRSSSTRSGGDLTLGLEPS 1066
QY 1073 EEEAPRSLAPSEAGSDVDFDGLGMGAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDG 1132
Db 1067 EEEAPRSLAPSEAGSDVDFDGLGMGAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDG 1126
QY 1133 YVAPLTCSPQPEYVNOPVDRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKOVFA 1192
Db 1127 YVAPLTCSPQPEYVNOPVDRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKOVFA 1186
QY 1193 FGGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLYWQDPPERGAPSTFKGTPTAENP 1252
Db 1187 FGGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLYWQDPPERGAPSTFKGTPTAENP 1246
QY 1253 EYLGLDVVPV 1261
Db 1247 EYLGLDVVPV 1255

RESULT 5
AAW01111
ID AAW01111 standard; Protein; 1255 AA.
XX AC AAW01111;
XX DT 01-JAN-1997 (first entry)
XX DE HER-2/neu protein.
XX KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
XX KW breast cancer; ovary cancer; colon cancer; lung cancer;
XX KW prostate cancer; immunisation; tumour; vaccine; vector.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 676..1255
XX FT /label= intracellular_domain
XX FT /note= "claimed domain, useful for immunisation"
XX PN W09630514-A1.
XX PD 03-OCT-1996.
XX
```

```
PF 28-MAR-1996; 96WO-US01689.
PR 31-MAR-1995; 95US-0414417.
PA (UNIW ) UNIV WASHINGTON.
XX Cheever MA, Disis ML;
XX WPI, 1996-455361/45.
XX N-PSDB; AAT40739.
XX DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
XX treatment of malignancies with which the HER-2/neu oncogene is
XX associated
XX Claim 2; Page 56-61; 71pp; English.
XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
XX the product of the HER-2/neu oncogene (see also AAT40739). The
XX protein is over-expressed in various cancers, including breast,
XX ovarian, colon, lung and prostate. The intracellular domain of the
XX protein can be used to immunise an animal against a malignancy with
XX which the oncogene is associated. The polypeptide can be produced
XX in transformed host cells for use in immunisation. Alternatively,
XX animal cells are transfected in vivo or ex vivo with a viral vector
XX that directs expression of the polypeptide.
XX Sequence 1255 AA;
Query Match 97.9%; Score 6684; DB 17; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;
QY 1 MELAALCRWGLLLALLPGGAASCTGCTDMKRLPASPETHLDMRLHLYGQCQVVOGNL 60
Db 1 MELAALCRWGLLLALLPGGAASCTGCTDMKRLPASPETHLDMRLHLYGQCQVVOGNL 60
QY 61 ELTYLPTNASLSFLQDIEQVGYVLIHNVQVPLQRLRIVRGTLQFEDNVALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIEQVGYVLIHNVQVPLQRLRIVRGTLQFEDNVALAVLNG 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
QY 181 LTLIDTNRSRACHPCSPCKGRCWGSSEDCQSLTRTVAGGCARCKGPIPFNNPTVSF 240
Db 181 LTLIDTNRSRACHPCSPCKGRCWGSSEDCQSLTRTVAGGCARCKGPIPFNNPTVSF 240
QY 241 WLRVPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPGHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 286
QY 293 GRYTFGASCVTACPNYVLTSDVSGCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLG 352
Db 287 GRYTFGASCVTACPNYVLTSDVSGCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLG 346
QY 353 MEHLREVRVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLBEI 412
Db 347 MEHLREVRVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLBEI 406
QY 413 TGYLYISAWPDSLPLDSVFNQVIRGRILHNGAYSLTLOGLGISWGLRSLRELGSGLA 472
Db 407 TGYLYISAWPDSLPLDSVFNQVIRGRILHNGAYSLTLOGLGISWGLRSLRELGSGLA 466
QY 473 LIHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCWGPPT 532
Db 467 LIHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCWGPPT 526
QY 533 QCVCNSQFLRGQECVBEICRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADQVFA 592
Db 527 QCVCNSQFLRGQECVBEICRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADQVFA 586
```

```
QY 593 CAHYKDPFCVARGSGVKPDLSPYMWKFPDEEGACQPCPINCTHSCVDLDDKGCPSAQ 652
Db 587 CAHYKDPFCVARGSGVKPDLSPYMWKFPDEEGACQPCPINCTHSCVDLDDKGCPSAQ 646
QY 653 RASPLTSISAVVGVLLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAM 712
Db 647 RASPLTSISAVVGVLLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAM 706
QY 713 PNQAOHRIKTELARKVKVLSGAFGTGYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 772
Db 707 PNQAOHRIKTELARKVKVLSGAFGTGYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 766
QY 773 ILDEAYVMAGVSPVYSRLGLCLSTVOLTPQMPYGCCLDHRNRLGSDQLLNCW 832
Db 767 ILDEAYVMAGVSPVYSRLGLCLSTVOLTPQMPYGCCLDHRNRLGSDQLLNCW 826
QY 833 MQIAKMSYLEDLVRLHRLDAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVP 892
Db 827 MQIAKMSYLEDLVRLHRLDAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVP 886
QY 893 KMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLLEKGERLPQPP 952
Db 887 KMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLLEKGERLPQPP 946
QY 953 CTIDVYMIWVKCWMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYR 1012
Db 947 CTIDVYMIWVKCWMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYR 1006
QY 1013 SILEDDMDGLVDABEYLVPQGGFCPPDPAPGAGGMVHRRHSSSTRSGGDLTLGLEPS 1072
Db 1007 SILEDDMDGLVDABEYLVPQGGFCPPDPAPGAGGMVHRRHSSSTRSGGDLTLGLEPS 1066
QY 1073 EEEAPRSLAPSEGAGSDVFDGLGMAAGKGLQSLPHTDPSPLQRYSEDPVPLPSETDG 1132
Db 1067 EEEAPRSLAPSEGAGSDVFDGLGMAAGKGLQSLPHTDPSPLQRYSEDPVPLPSETDG 1126
QY 1133 YVAPLTCSPQPEYVNPQDVRRPQSPREGPLPAARPAATLERAKTLPSPKNGVVKDVA 1192
Db 1127 YVAPLTCSPQPEYVNPQDVRRPQSPREGPLPAARPAATLERAKTLPSPKNGVVKDVA 1186
QY 1193 FGGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENP 1252
Db 1187 FGGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENP 1246
QY 1253 EYLGIDVAV 1261
Db 1247 EYLGIDVAV 1255

RESULT 6
AAW92406
ID AAW92406 standard; Protein; 1255 AA.
XX
AC AAW92406;
XX
DT 21-APR-1999 (first entry)
XX
DE Human HER-2/neu oncogene protein.
XX
KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 676..1255
FT /note= "region which elicits immune response"
XX
PN US5869445-A.
XX
PD 09-FEB-1999.
XX
PF 01-APR-1996; 96US-0625101.
```

```
XX
PR 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
PA (UNI* ) UNIV WASHINGTON.
XX
PI Cheever MA, Disis ML;
XX
DR WPI; 1999-152835/13.
DR N-PSDB; AAX01912.
XX
PT Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX
PS Claim 3; Column 31-38; 26pp; English.
XX
CC This sequence represents the human HER-2/neu oncogene protein. A fragment
CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ Sequence 1255 AA;

Query Match 97.9%; Score 6684; DB 20; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;

QY 1 MELAALCAKRWGLLLALLPPGAASSTQVCTGDMKRLPASPETHLDMLRHLYQSCVVOGNL 60
Db 1 MELAALCAKRWGLLLALLPPGAASSTQVCTGDMKRLPASPETHLDMLRHLYQSCVVOGNL 60
QY 61 ELTYLPTNASLSFLQDIOEQVGYVLIHNVQRVPLQRLIRVRGTQOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEQVGYVLIHNVQRVPLQRLIRVRGTQOLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTGASPGGLRELQRLSRLTEILKGVLIQRNPOLCYQDTILWKDIFHKNQLA 180
Db 121 DPLNNTTPTGASPGGLRELQRLSRLTEILKGVLIQRNPOLCYQDTILWKDIFHKNQLA 180
QY 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTVTCAGGCARCKGPLPNNFTVSF 240
Db 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTVTCAGGCARCKGPLPNNFTVSF 240
QY 241 WLRVPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 292
Db 233 -----TDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 286
QY 293 GRVTFGASCVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLG 352
Db 287 GRVTFGASCVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLG 346
QY 353 MEHLREVRVTSANIOEPAGCKIFGSLAFIPESPDGDPASNTAPLQPEOLOVFTLEBI 412
Db 347 MEHLREVRVTSANIOEPAGCKIFGSLAFIPESPDGDPASNTAPLQPEOLOVFTLEBI 406
QY 413 TGYLYISAWPDSLPLDSVFNQLQVIRGRIHNGAYSILTQGLIGISWLGSLRLSLGSLA 472
Db 407 TGYLYISAWPDSLPLDSVFNQLQVIRGRIHNGAYSILTQGLIGISWLGSLRLSLGSLA 466
QY 473 LIHNTHLCFVHTVPWDLFRNPHQALLHTANRPEDECVGEGGLACHQI-CARGHCWGPGPT 532
Db 467 LIHNTHLCFVHTVPWDLFRNPHQALLHTANRPEDECVGEGGLACHQI-CARGHCWGPGPT 526
QY 533 QCVNCSOFLRGQECVBEICRVLOGLPREYVNAHCLPCHPECPQNGSVTCTGPEADQV 592
Db 527 QCVNCSOFLRGQECVBEICRVLOGLPREYVNAHCLPCHPECPQNGSVTCTGPEADQV 586
```


Db 587 CAHYKDPFCVACPSGVFDPDLSYMPIMKFFDEGACQPCPINCTHSCVDLDDKGCPAEQ 646
Qy 653 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAM 712
Db 647 RASPLTSIIISAVVIGILLVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAM 706
Qy 713 PNOAQRILKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVIRENTSPKANKE 772
Db 707 PNOAQRILKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVIRENTSPKANKE 766
Qy 773 ILDEAVYVAGVSPYVSRLLGLICTSTVOLVTQLMYPGCLLDHVRNRLGSDLLNWC 832
Db 767 ILDEAVYVAGVSPYVSRLLGLICTSTVOLVTQLMYPGCLLDHVRNRLGSDLLNWC 826
Qy 833 MOIAKMSVLEDRVLRHDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVP 892
Db 827 MOIAKMSVLEDRVLRHDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVP 886
Qy 893 KMALESILRRRFTHOSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPI 952
Db 887 KMALESILRRRFTHOSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPI 946
Qy 953 CTIDVYIMVKCMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYR 1012
Db 947 CTIDVYIMVKCMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYR 1006
Qy 1013 SLEDDDDMDGLVDABEYLVPOQGFCDPDPAPGAGMVHRRSSSTRSGGDLTLGLEPS 1072
Db 1007 SLEDDDDMDGLVDABEYLVPOQGFCDPDPAPGAGMVHRRSSSTRSGGDLTLGLEPS 1066
Qy 1073 EEEAPRSLAPSEAGSDVFDGLGMAAGKQSLPHTDPSPLORYSEPTVPLPSETDG 1132
Db 1067 EEEAPRSLAPSEAGSDVFDGLGMAAGKQSLPHTDPSPLORYSEPTVPLPSETDG 1126
Qy 1133 YVAPLTCSQPBYVNVQPDVRPQPPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFA 1192
Db 1127 YVAPLTCSQPBYVNVQPDVRPQPPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFA 1186
Qy 1193 FCGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLYWDDODPPERGAPPTFKGTPTAENP 1252
Db 1187 FCGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLYWDDODPPERGAPPTFKGTPTAENP 1246
Qy 1253 EYLGLDVPV 1261
Db 1247 EYLGLDVPV 1255

RESULT 8

AAy84780
ID AAY84780 standard; Protein; 1255 AA.
XX
AC AAY84780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
XX wound healing.
XX
OS Homo sapiens.
XX
PN WO200020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX

PA (UYMC-) UNIV MCMASTER.
XX
PI Muller WJ, Siegel PM;
XX
DR WPI: 2000-303768/26.
XX
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPLICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;
Query Match 97.9%; Score 6684; DB 21; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;
Qy 1 MELAALCRWGLLALLPFGAASQVCTGDMKRLRPASPETHLDMRLHLYQGCQVVOGNL 60
Db 1 MELAALCRWGLLALLPFGAASQVCTGDMKRLRPASPETHLDMRLHLYQGCQVVOGNL 60
Qy 61 ELTYLPTNASLFLQDIOEVQGYVLIANQVRQVPLQRLRIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIOEVQGYVLIANQVRQVPLQRLRIRVGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGASPGGLRELQRLSITELKGGVLTORNPOLCYQDTILWKDIFHKQNOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSITELKGGVLTORNPOLCYQDTILWKDIFHKQNOLA 180
Qy 181 LTLIDNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGACRCKPLPNNFTVSF 240
Db 181 LTLIDNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGACRCKPLPNNFTVSF 232
Qy 241 WLRVPKVSASHLE-----KHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPE 292
Db 233 -----TDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDTFESMPNPE 286
Qy 293 GRYTFGASCVTACPYNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLG 352
Db 287 GRYTFGASCVTACPYNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLG 346
Qy 353 MEHLREVRVTSANTQEFAGCKKIFGSLAFIPESPDGDPASNTAPLQEPQLQVFTLESI 412
Db 347 MEHLREVRVTSANTQEFAGCKKIFGSLAFIPESPDGDPASNTAPLQEPQLQVFTLESI 406
Qy 413 TGLYLISAWPDSLPDLSVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSURELSGLA 472
Db 407 TGLYLISAWPDSLPDLSVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSURELSGLA 466
Qy 473 LIHNTHLCFVHTVPWDOLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCGPGPT 532
Db 467 LIHNTHLCFVHTVPWDOLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCGPGPT 526
Qy 533 QCVNCSQFLRGQECVVECRVLQGLPREYVNAHCLPCHPECPQNGSVTFCFGEADQCV 592
Db 527 QCVNCSQFLRGQECVVECRVLQGLPREYVNAHCLPCHPECPQNGSVTFCFGEADQCV 586
Qy 593 CAHYKDPFCVACPSGVKPDLSYMPIMKFFDEGACQPCPINCTHSCVDLDDKGCPAEQ 652

Db 587 CAHYKDPFCVACPSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEQ 646
Qy 653 RASPLTSIISAVVGIILVVVGLVGVVGLIKRROOKIRKYTMRLLOETELVEPLTSGAM 712
Db 647 RASPLTSIISAVVGIILVVVGLVGVVGLIKRROOKIRKYTMRLLOETELVEPLTSGAM 706
Qy 713 PNOAQMRILKETELRKVKVGLGSAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 772
Db 707 PNOAQMRILKETELRKVKVGLGSAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 766
Qy 773 ILDEAVYMACVGSPPYVSRLLGICLTSTVQLVTLQMPYGCULLDVHVRNRLGSDLLNWC 832
Db 767 ILDEAVYMACVGSPPYVSRLLGICLTSTVQLVTLQMPYGCULLDVHVRNRLGSDLLNWC 826
Qy 833 MQIAKMSYLEDLVRLVHRDLAARNLVKSPNHVKITDFGLARLLDDDETYHADGGKVP 892
Db 827 MQIAKMSYLEDLVRLVHRDLAARNLVKSPNHVKITDFGLARLLDDDETYHADGGKVP 886
Qy 893 KMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 952
Db 887 KMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 946
Qy 953 CTIDVYIMVKCWMIDSECRPRFRELVSFERSMARDPQRFVITQNEDLGPASPLDSTFYR 1012
Db 947 CTIDVYIMVKCWMIDSECRPRFRELVSFERSMARDPQRFVITQNEDLGPASPLDSTFYR 1006
Qy 1013 SLEDDDDMGDLVDAEYLVPOQGFCCPDAPAGGVMVHRRHRSSTRSGGDLTLGLEPS 1072
Db 1007 SLEDDDDMGDLVDAEYLVPOQGFCCPDAPAGGVMVHRRHRSSTRSGGDLTLGLEPS 1066
Qy 1073 EEBAPRSPLAPSGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDG 1132
Db 1067 EEBAPRSPLAPSGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDG 1126
Qy 1133 YVAPLTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERPRTLSPKNGVYKDVFA 1192
Db 1127 YVAPLTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERPRTLSPKNGVYKDVFA 1186
Qy 1193 FGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENP 1252
Db 1187 FGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENP 1246
Qy 1253 EYLGLDVVP 1261
Db 1247 EYLGLDVVP 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX AAB85458;
AC
XX
XX 25-SEP-2001 (first entry)
DT
DE Human HER-2/neu protein.
XX
XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
XX Homo sapiens.
XX
XX WO200153463-A2.
XX
XX 26-JUL-2001.
XX
XX 19-JAN-2001; 2001WO-US01850.
PF
XX
XX 21-JAN-2000; 2000US-0177545.
PR
XX
XX (CORI-) CORIXA CORP.
PA
XX

PI Cheever MA, Hand-Zimmermann S;
XX WPI; 2001-476112/51.
DR N-PSDB; AAH23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer
XX
PS Claim 2; Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;

Query Match 97.9%; Score 6684; DB 22; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;
Qy 1 MELAAALCRWGLLALLPPGNAASQVCTGTDMLKRLPASPTHLDMLRHLVYGCQVVOGNL 60
Db 1 MELAAALCRWGLLALLPPGNAASQVCTGTDMLKRLPASPTHLDMLRHLVYGCQVVOGNL 60
Qy 61 ELTYLPTNASLSPLQDIQEVQGVVIAHNOVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSPLQDIQEVQGVVIAHNOVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVGTASPGGLREQLRSLTEILKGGVLIORNPOLCYQDTTLTKWDIFHKNOLA 180
Db 121 DPLNNTPTVGTASPGGLREQLRSLTEILKGGVLIORNPOLCYQDTTLTKWDIFHKNOLA 180
Qy 181 LTLIDNRSRACHPCSPMKGSCWGESSEDCOSLRTVTCAGGCARCKGPLPFNNFTVSF 240
Db 181 LTLIDNRSRACHPCSPMKGSCWGESSEDCOSLRTVTCAGGCARCKGPLPFNNFTVSF 232
Qy 241 WLVRPKVYASHLE-----KHSDCLACILFHNHSGICELHCPALVYNTDTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACILFHNHSGICELHCPALVYNTDTFESMPNPE 286
Qy 293 GRYTFGASCVTACPYNYLSTDVGSCTLVCPHLHNOEVTAEQGTQCEKCKSPCARVCYGLG 352
Db 287 GRYTFGASCVTACPYNYLSTDVGSCTLVCPHLHNOEVTAEQGTQCEKCKSPCARVCYGLG 346
Qy 353 MEHLREVRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPELLEE 412
Db 347 MEHLREVRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPELLEE 406
Qy 413 TGYLYISAWPDSLPLDSVFONQVIRGRILHNGAYSLTLQGLIGISWGLSLRSLGSLA 472
Db 407 TGYLYISAWPDSLPLDSVFONQVIRGRILHNGAYSLTLQGLIGISWGLSLRSLGSLA 466
Qy 473 LIHNTHLCFVHTVPMWDLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGPT 532
Db 467 LIHNTHLCFVHTVPMWDLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGPT 526
Qy 533 QCVNCSQFLRGQECVCECRVLQGLPREYVNAHCLPCHPECOQNGSVTCFGEADOCVA 592
Db 527 QCVNCSQFLRGQECVCECRVLQGLPREYVNAHCLPCHPECOQNGSVTCFGEADOCVA 586
Qy 593 CAHYKDPFCVACPSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEQ 652
Db 587 CAHYKDPFCVACPSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEQ 646
Qy 653 RASPLTSIISAVVGIILVVVGLVGVVGLIKRROOKIRKYTMRLLOETELVEPLTSGAM 712

Db 647 RASPLTSIIISAVGILLVVVGVVFGIILIKRRQKIRKYTWRRLLQETELVEPLTPSGAM 706
Qy 713 PNOQMRLKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKE 772
Db 707 PNOQMRLKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKE 766
Qy 773 ILDEAYVMAGVSGPVSRLLGICLTSTVOLVTQLMPYGCCLLDHVRENRGLGSODLLNWC 832
Db 767 ILDEAYVMAGVSGPVSRLLGICLTSTVOLVTQLMPYGCCLLDHVRENRGLGSODLLNWC 826
Qy 833 MOIAGKSYLEDVRLVHDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGKVP 892
Db 827 MOIAGKSYLEDVRLVHDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGKVP 886
Qy 893 KMALESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 952
Db 887 KMALESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 946
Qy 953 CTIDVYIMVWKWIDSECRPRFRELVSFMRMADPQRFVVIQNEIDLGPASPLDSTFYR 1012
Db 947 CTIDVYIMVWKWIDSECRPRFRELVSFMRMADPQRFVVIQNEIDLGPASPLDSTFYR 1006
Qy 1013 SLLEDDMGDLVDAAEYLVPOQGFPCPDPAAGGMVHRRSSSTRSGGDLTLGLEPS 1072
Db 1007 SLLEDDMGDLVDAAEYLVPOQGFPCPDPAAGGMVHRRSSSTRSGGDLTLGLEPS 1066
Qy 1073 EEEAPRSLAPSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDG 1132
Db 1067 EEEAPRSLAPSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDG 1126
Qy 1133 YVAPLTCSPQPEYVNPQDVRPQPSREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVA 1192
Db 1127 YVAPLTCSPQPEYVNPQDVRPQPSREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVA 1186
Qy 1193 FGGAVENPEYLTPOGGAAPHPPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENP 1252
Db 1187 FGGAVENPEYLTPOGGAAPHPPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENP 1246
Qy 1253 EYLGDLVPV 1261
Db 1247 EYLGDLVPV 1255

RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
OS Homo sapiens.
XX
XX WO200141787-A1.
XX
XX 14-JUN-2001.
XX
XX 11-DEC-2000; 2000WO-US33591.
XX
XX 10-DEC-1999; 99US-0458299.
XX
XX (EPIM-) EPIMUNE INC.
XX
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
XX WPI; 2001-374995/39.

XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX Disclosure; Page 15; 199pp; English.
XX
XX The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II), comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
XX Sequence 1255 AA;

Query Match 97.9%; Score 6684; DB 22; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;
Qy 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMLRLPASPETHLDMRLHYQCCVQVQGNL 60
Db 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMLRLPASPETHLDMRLHYQCCVQVQGNL 60
Qy 61 ELTYLPTNASLFLQDIEQVQYVLIHNOVQVPLQRLRIVRGTLQEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIEQVQYVLIHNOVQVPLQRLRIVRGTLQEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELOQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELOQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGRCWSESSEDCQSLTRTVCCAGGCARCKGPLPFPNFTVSF 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWSESSEDCQSLTRTVCCAGGCARCKGPLPFPNFTVSF 240
Qy 241 WLVRPKVSASHLE-----KHSDCLACLHFNHSGICELHCPALVTNTDFTFEMPNE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTNTDFTFEMPNE 286
Qy 293 GRVTFGASCVTACPNYVLSTDVSGCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLG 352
Db 287 GRVTFGASCVTACPNYVLSTDVSGCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLG 346
Qy 353 MEHLREVRVAVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEBI 412
Db 347 MEHLREVRVAVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEBI 406
Qy 413 TGYLYISAWPDSLPLSLVFNQNLQVIRGRILHNGAYSILTQGLGSLWLSRLSRELSGLA 472
Db 407 TGYLYISAWPDSLPLSLVFNQNLQVIRGRILHNGAYSILTQGLGSLWLSRLSRELSGLA 466
Qy 473 LIHNHTLCLCFVHTVPWDQLFRNPHQALHTANRDECEVCGELGACHQLCARGHCWCPGPT 532

| | | | |
|-----------|---|--|------|
| Db | 527 | QCNCVQFLRGQCEVECRVLOGLPREYNARHCLPCHPECOQPNGSVTCFGFEADQCUA | 586 |
| Qy | 593 | CAHYKDPFFCVCARCPGKVPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKCPAEQ | 652 |
| Db | 587 | CAHYKDPFFCVCARCPGKVPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKCPAEQ | 646 |
| Qy | 653 | RASPLTSIVSAVVGILLVVLGVVGGIILKRRQOKIRKYTHRLLQETELVEPLTPSGAM | 712 |
| Db | 647 | RASPLTSIISAVVGILLVVLGVVGGIILKRRQOKIRKYTHRLLQETELVEPLTPSGAM | 706 |
| Qy | 713 | PNQAQMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKE | 772 |
| Db | 707 | PNQAQMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKE | 766 |
| Qy | 773 | ILDEAYNAGVCSPIVSRLLGLICLTSTVOLVTQLMPYGCGLLDHVHRENRGLSQDLLNWC | 832 |
| Db | 767 | ILDEAYNAGVCSPIVSRLLGLICLTSTVOLVTQLMPYGCGLLDHVHRENRGLSQDLLNWC | 826 |
| Qy | 833 | MOIAGMSYLEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGKVP | 892 |
| Db | 827 | MOIAGMSYLEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGKVP | 886 |
| Qy | 893 | KWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI | 952 |
| Db | 887 | KWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI | 946 |
| Qy | 953 | CTIDVYMIWVKWIMIDSECRFRFRELVSFSEFMRDQRFVVIQNEGLGPASPLDSTFYR | 1012 |
| Db | 947 | CTIDVYMIWVKWIMIDSECRFRFRELVSFSEFMRDQRFVVIQNEGLGPASPLDSTFYR | 1006 |
| Qy | 1013 | SLLEDMDGDLVDAEYILVPOQFFCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPS | 1072 |
| Db | 1007 | SLLEDMDGDLVDAEYILVPOQFFCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPS | 1066 |
| Qy | 1073 | EEEAAPRSLAPSEAGSDVDFDGLGMAAKGLQSLPTHDPSPLOYSDDPTVPLPSETDG | 1132 |
| Db | 1067 | EEEAAPRSLAPSEAGSDVDFDGLGMAAKGLQSLPTHDPSPLOYSDDPTVPLPSETDG | 1126 |
| Qy | 1133 | YVAPLTCSPQPEYNQPDVVRQPPSPREGPLPAARPAAGATLERAKTLPFGKNGVVKDVA | 1192 |
| Db | 1127 | YVAPLTCSPQPEYNQPDVVRQPPSPREGPLPAARPAAGATLERAKTLPFGKNGVVKDVA | 1186 |
| Qy | 1193 | FGGAIVENPEYLTPOGGAAPQHPHPPAFSPADNLVYWDODPPERGAAPPSTFKGTAEHP | 1252 |
| Db | 1187 | FGGAIVENPEYLTPOGGAAPQHPHPPAFSPADNLVYWDODPPERGAAPPSTFKGTAEHP | 1246 |
| Qy | 1253 | EYLGLDVPV 1261 | |
| Db | 1247 | EYLGLDVPV 1255 | |
| RESULT 12 | | | |
| AAE20479 | | | |
| ID | AAE20479 | standard; Protein; 1255 AA. | |
| AC | AAE20479; | | |
| AC | AAE20479; | | |
| DT | 01-JUL-2002 | (first entry) | |
| DE | Human Her-2/neu protein. | | |
| KW | Human; Her-2/Neu protein; immune response; gene therapy; breast cancer; | | |
| KW | human leukocyte antigen; HLA; vaccine; malignancy; cytostatic. | | |
| OS | Homo sapiens. | | |
| FX | Key | Location/Qualifiers | |
| FT | Region | 1021..1030 | |
| FT | | /note= "Naturally processed HLA-B44-restricted epitope" | |
| PN | WO200214503-A2. | | |
| XX | | | |

| | | | |
|----|---|--|-----|
| PD | 21-FEB-2002. | | |
| XX | 14-AUG-2001; 2001WO-US41733. | | |
| PF | 14-AUG-2000; 2000US-225152P. | | |
| PR | 28-SEP-2000; 2000US-236428P. | | |
| PR | 21-FEB-2001; 2001US-270520P. | | |
| XX | (CORI-) CORIXA CORP. | | |
| XX | Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD, McNeill PD, Vedwick TS, WPI; 2002-280758/32. N-PSDB; AAD32743. | | |
| DR | Novel isolated Her-2/Neu polypeptide composition useful for therapy, prevention and diagnosis of cancer, preferably breast cancer | | |
| DR | Disclosure; Page 114-117; 129pp; English. | | |
| XX | The invention relates to an isolated Her-2/Neu polypeptide composition effective for eliciting an immune response. The invention is useful for eliciting an immune response in a patient, where the patient is human leukocyte antigen (HLA)-B44 positive or is affected with breast cancer. The composition is useful for the therapy and diagnosis of cancer, preferably breast cancer, in pharmaceutical compositions, e.g., vaccine and other compositions for the diagnosis, prevention and treatment of human malignancies, for stimulating and/or expanding T cells specific for Her-2/Neu polypeptide and for inhibiting the development of cancer in a patient. The invention is useful for stimulating a T cell response in a human patient, as probe or primer for nucleic acid hybridisation, to selectively form duplex molecules with complementary stretches of the entire Her-2/Neu gene or gene fragments of interest, to isolate a full length gene from a suitable library, and to direct expression of a polypeptide in appropriate host cells. The composition is useful in prophylactic or therapeutic applications and for the treatment of cancer, preferably for the immunotherapy of breast cancer and other Her-2/Neu-associated malignancies. The invention is useful in gene therapy. The present sequence is human Her-2/neu protein. | | |
| XX | Sequence 1255 AA; | | |
| SQ | Query Match 97.9%; Score 6684; DB 23; Length 1255; Best Local Similarity 97.6%; Pred. No. 0; Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2; | | |
| Qy | 1 | MELAAALCPWGLLLALLPPGAASCTCTGDMKRLPASPETHLDMRLHYQCVVQGNL | 60 |
| Db | 1 | MELAAALCPWGLLLALLPPGAASCTCTGDMKRLPASPETHLDMRLHYQCVVQGNL | 60 |
| Qy | 61 | ELTYLPTNASLSFLQDIOEVQGYVLIHNRQVPLQRLRIVRGTLQFEDNYALAVLDNG | 120 |
| Db | 61 | ELTYLPTNASLSFLQDIOEVQGYVLIHNRQVPLQRLRIVRGTLQFEDNYALAVLDNG | 120 |
| Qy | 121 | DPLNNTTPTVTCASPGGLREQLRLSLEILKGVLIQRNPQLCYQDTILWKDIFHKNNQLA | 180 |
| Db | 121 | DPLNNTTPTVTCASPGGLREQLRLSLEILKGVLIQRNPQLCYQDTILWKDIFHKNNQLA | 180 |
| Qy | 181 | LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRVCAGGCARCKGLPFFNFTVSF | 240 |
| Db | 181 | LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRVCAGGCARCKGLPFFNFTVSF | 240 |
| Qy | 241 | WLRVPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVTNTDTFESMPNPE | 292 |
| Db | 233 | -----TDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTNTDTFESMPNPE | 286 |
| Qy | 293 | GRYTFGASCVTACPYNYLSTDVGSCITLVCPPLHNEVTAEDGTQRCCKSKPCARVCYGLG | 352 |
| Db | 287 | GRYTFGASCVTACPYNYLSTDVGSCITLVCPPLHNEVTAEDGTQRCCKSKPCARVCYGLG | 346 |
| Qy | 353 | MEHLREVRANTSANIQEPAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLOVFTLBEI | 412 |

Db 347 MEHLREAVTSANIQEFAGCKKIFGLAPLPESEFDGDPASNTAPLQPEQLQVFETLEBI 406
Qy 413 TGYLYISAWPDSLPDLSVFQNLQVIRGRILHNGAVSLTLQGLGISWLGSLRSLRELGSGLA 472
Db 407 TGYLYISAWPDSLPDLSVFQNLQVIRGRILHNGAVSLTLQGLGISWLGSLRSLRELGSGLA 466
Qy 473 LIHNTHLCFVHTVPMQDLFRNPHQALLHTANRPEDECVGEGACHQLCARGHMGPGPT 532
Db 467 LIHNTHLCFVHTVPMQDLFRNPHQALLHTANRPEDECVGEGACHQLCARGHMGPGPT 526
Qy 533 QCVNCSQFLRGQECVEECRVLOGLPREYNARHCLPCHPECPQNGSVTCFGEADQCYA 592
Db 527 QCVNCSQFLRGQECVEECRVLOGLPREYNARHCLPCHPECPQNGSVTCFGEADQCYA 586
Qy 593 CAHYKDPFPCVACRPSGVKPDLSYMPIWKFPEEGACQPCINCHTHSCVDLDDKGCPEAQ 652
Db 587 CAHYKDPFPCVACRPSGVKPDLSYMPIWKFPEEGACQPCINCHTHSCVDLDDKGCPEAQ 646
Qy 653 RASPLTSIYSAVVGILLVVLVGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 712
Db 647 RASPLTSIYSAVVGILLVVLVGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 706
Qy 713 PNOAQRILKETELRKVKVLGSGAFCTVYKGIWIPGENVKIPVAIKVIRENTSPKANKE 772
Db 707 PNOAQRILKETELRKVKVLGSGAFCTVYKGIWIPGENVKIPVAIKVIRENTSPKANKE 766
Qy 773 ILDEAYVMAGVSPYVRLIGLICLTSTVOLVTOLMPYGCLLDHVRENRLGSGODLLNWC 832
Db 767 ILDEAYVMAGVSPYVRLIGLICLTSTVOLVTOLMPYGCLLDHVRENRLGSGODLLNWC 826
Qy 833 MQIAKMSYLEVDRLVHRDLARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPI 892
Db 827 MQIAKMSYLEVDRLVHRDLARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPI 886
Qy 893 KMALESILRRRFTHOSDVWSYGVTVWELMTFCAPYDGI PAREIPDLLEKGERLPQPI 952
Db 887 KMALESILRRRFTHOSDVWSYGVTVWELMTFCAPYDGI PAREIPDLLEKGERLPQPI 946
Qy 953 CTIDVYIMVWKWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEGLGASPLDSTFYR 1012
Db 947 CTIDVYIMVWKWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEGLGASPLDSTFYR 1006
Qy 1013 SLEDDMDGLVDAAEYLYVPOQFFCPDPAPGAGVHHRSSSTRSGGDLTLGLEPS 1072
Db 1007 SLEDDMDGLVDAAEYLYVPOQFFCPDPAPGAGVHHRSSSTRSGGDLTLGLEPS 1066
Qy 1073 EEEAPSPAPSEAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSETDG 1132
Db 1067 EEEAPSPAPSEAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSETDG 1126
Qy 1133 YVAPLTCSQPEYVNPQDVVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFA 1192
Db 1127 YVAPLTCSQPEYVNPQDVVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFA 1186
Qy 1193 FCGAVENPEYLTPOGGAAPQPPHPPAFSPAFNLYYWDQDPPERGAPPSTFKGTPTAENP 1252
Db 1187 FCGAVENPEYLTPOGGAAPQPPHPPAFSPAFNLYYWDQDPPERGAPPSTFKGTPTAENP 1246
Qy 1253 EYGLGDVPV 1261
Db 1247 EYGLGDVPV 1255

RESULT 13

AAMS1143

ID AAMS1143 standard; Protein; 1255 AA.

XX

AC AAMS1143;

XX

DT 17-JUN-2002 (first entry)

XX

DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX

Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
tyrosine kinase; receptor; c-erbB2; gene therapy.

Homo sapiens.

Key Location/Qualifiers
Domain 1..653
/note= "extracellular domain"
Domain 676..1255
/note= "intracellular domain"
Domain 990..1255
/note= "phosphorylation domain"

W0200212341-A2.

14-FEB-2002.

03-AUG-2001; 2001WO-US24283.

03-AUG-2000; 2000US-0632507.

(CORI-) CORIXA CORP.

(SMIX) SMITHKLINE BEECHAM BIOLOGICALS.

Cheever MA, Cheysen D;

WPI: 2002-241743/29.

N-PSDB; ABA92250.

Her-2/neu fusion protein for treating or preventing cancer by eliciting
or enhancing an immune response to the protein, has Her-2/neu
extracellular domain fused to Her-2/neu intracellular or
phosphorylation domain -

Claim 68; Fig 7; 14pp; English.

The present sequence is that of human Her-2/neu (p185 glycoprotein
or c-erbB2), an oncogenic self-protein and target for anti-cancer
vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
in a variety of cancers, including breast, ovarian, colon, lung and
prostate cancer. Her-2/neu is a member of the tyrosine kinase
family of receptor-like glycoproteins. It comprises an extracellular
domain with homology to the epidermal growth factor receptor
(EGFR), a highly hydrophobic transmembrane domain and a C-terminal
intracellular domain that also shows homology to EGFR. Its
overexpression correlates with a poor prognosis in breast and
ovarian cancers. The invention provides Her-2/neu fusion
proteins, nucleic acids encoding them, viral vectors, and vaccines
comprising the fusion proteins or nucleic acid molecules. In
preferred fusion proteins, the extracellular domain of a Her-2/neu
protein is fused to a Her-2/neu intracellular domain or
phosphorylation domain (or its DeltaAPD fragment). An immune
response to Her-2/neu protein is elicited or enhanced by
administering the fusion protein in the form of a vaccine, or by
transfecting cells of an animal ex vivo with a nucleic acid
encoding the fusion protein, and delivering the transfected cells
to the animal. The fusion proteins, nucleic acids, and isolated
specific T-cells are useful for inhibiting the development of a
cancer, especially breast, ovarian, colon, lung or prostate cancer
in a patient. T cells that specifically react with a Her-2/neu
fusion protein can be used to remove tumour cells from a sample in
order to inhibit the development of cancer in a patient.

Sequence 1255 AA;

Query Match 97.9%; Score 6684; DB 23; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHLMLRHLYQSCVVGQNL 60

Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHLMLRHLYQSCVVGQNL 60

```
QY 61 ELTYLPTNASLSFLQDIEQVGYVLI AHNVQVFLORLRI VRTQQLFEDNYALAVLNG 120
DB 61 ELTYLPTNASLSFLQDIEQVGYVLI AHNVQVFLORLRI VRTQQLFEDNYALAVLNG 120
QY 121 DPLNNTTPTGASPGGLRELQRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKQNOLA 180
DB 121 DPLNNTTPTGASPGGLRELQRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKQNOLA 180
QY 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDQSLTRTVACGACRCKGPLPNNFTVSF 240
DB 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDQSLTRTVACGACRCKGPLPNNFTVSF 240
QY 241 WLRVPKVSASHLE-----KHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPE 292
DB 241 WLRVPKVSASHLE-----KHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPE 292
QY 293 GRYTEGASCVTACPNYLSLTDVGSCTLCVPLHNEVTAEDGTQRCCKSKPCARVCYGLG 352
DB 293 GRYTEGASCVTACPNYLSLTDVGSCTLCVPLHNEVTAEDGTQRCCKSKPCARVCYGLG 352
QY 353 MEHLREVRVTSANTQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEBI 412
DB 353 MEHLREVRVTSANTQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEBI 412
QY 413 TGYLYISAWPDSLPDLSVFNQNLQVIRGRILHNGAYSLTLOGLGISWGLRLSRELGSGLA 472
DB 413 TGYLYISAWPDSLPDLSVFNQNLQVIRGRILHNGAYSLTLOGLGISWGLRLSRELGSGLA 472
QY 473 LIHNTHLCLFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPT 532
DB 473 LIHNTHLCLFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPT 532
QY 533 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPEQOPQNGSVTCFGEADQCV 592
DB 533 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPEQOPQNGSVTCFGEADQCV 592
QY 593 CAHYKDPFCVACRCSGKVPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKCPAEQ 652
DB 593 CAHYKDPFCVACRCSGKVPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKCPAEQ 652
QY 653 RASPLTISVSAVVGILLVVLGVVFGILIKRRQKIRKYTWRRLLQETELVEPLTPSGAM 712
DB 653 RASPLTISVSAVVGILLVVLGVVFGILIKRRQKIRKYTWRRLLQETELVEPLTPSGAM 712
QY 713 PNOQMRLKETELRKVKVLSGARGTVYKGIWIPDGENVKIPVAIKVLRNRENTSPKANKE 772
DB 713 PNOQMRLKETELRKVKVLSGARGTVYKGIWIPDGENVKIPVAIKVLRNRENTSPKANKE 772
QY 773 ILDEAYVMAGVSPVSRLLGICLTSTVOLVTOLMPYGLLDHVRNENRGLGSQDLNWC 832
DB 773 ILDEAYVMAGVSPVSRLLGICLTSTVOLVTOLMPYGLLDHVRNENRGLGSQDLNWC 832
QY 833 MQIAGMSYLEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGCKVPI 892
DB 833 MQIAGMSYLEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGCKVPI 892
QY 893 KMALESILRRFRTHQSDVNSVGYTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 952
DB 893 KMALESILRRFRTHQSDVNSVGYTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 952
QY 953 CTIDVYMIWKCWIMDSERPRFRELVSFESRMARDPQRFVVIQNEIDLGASPLDSTFYR 1012
DB 953 CTIDVYMIWKCWIMDSERPRFRELVSFESRMARDPQRFVVIQNEIDLGASPLDSTFYR 1012
QY 1013 SLEDDDMGDLVDAEYLVPOQGFCCPDPAFAGGVMVHRRSSSTRSGGDLTLGLBPS 1072
DB 1013 SLEDDDMGDLVDAEYLVPOQGFCCPDPAFAGGVMVHRRSSSTRSGGDLTLGLBPS 1072
QY 1073 EEEAPRSPLABSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLQRYSEDPTVPLPSETDG 1132
DB 1073 EEEAPRSPLABSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLQRYSEDPTVPLPSETDG 1132
QY 1133 YVAPLTCSPOPEYVNPQDVPRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVA 1192
```

```
DB 1127 YVAPLTCSPOPEYVNPQDVPRPQPPSPREGPLPAARPAGATLERPKTILSPGKNGVVKDVA 1186
QY 1193 FCGAVENPEYLTPOCGAARPOPHPPPAFSPAFNLYYWDQDPPERGAAPPSTFKGTPTAENP 1252
DB 1187 FCGAVENPEYLTPOCGAARPOPHPPPAFSPAFNLYYWDQDPPERGAAPPSTFKGTPTAENP 1246
QY 1253 EYLGLDVPV 1261
DB 1247 EYLGLDVPV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; Chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
XX
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
DR WPI; 2002-280741/32.
XX
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
PS Disclosure; Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 97.9%; Score 6684; DB 23; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;

QY 1 MELAALCRWGLLLALLPFGAASQTCTGDMKRLPASPEHMLRLHLYOGCQVQGNL 60
DB 1 MELAALCRWGLLLALLPFGAASQTCTGDMKRLPASPEHMLRLHLYOGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIEQVGYVLI AHNVQVFLORLRI VRTQQLFEDNYALAVLNG 120
```

Db 61 ELTYLPTNASLSFLQDQIEVQGVLIHQNVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGASPGSLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGASPGSLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKGSRGCGESSEDQSLTRTVCAGGCARCKGPLPFNNFTVSF 240
Db 181 LTLIDTNRSRACHPCSPMCKGSRGCGESSEDQSLTRTVCAGGCARCKGPLP----- 232
Qy 241 WLRVPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 286
Qy 293 GRYTFEGASCVTACPNYLSLTDVGSCTLVCPLNHNOEVTAEADGTORCEKSKPCARVCYGLG 352
Db 287 GRYTFEGASCVTACPNYLSLTDVGSCTLVCPLNHNOEVTAEADGTORCEKSKPCARVCYGLG 346
Qy 353 MEHLREVAVTSANIQEFAGCKKI FGS LAF L P ES FGD DPASNTAP LQPEQLQV FETLEBI 412
Db 347 MEHLREVAVTSANIQEFAGCKKI FGS LAF L P ES FGD DPASNTAP LQPEQLQV FETLEBI 406
Qy 413 TGYLYISAWPDSLPLDSVFQNLQVIRGRILHNGAYS LTLQGLGISWGLRSLRELGLA 472
Db 407 TGYLYISAWPDSLPLDSVFQNLQVIRGRILHNGAYS LTLQGLGISWGLRSLRELGLA 466
Qy 473 LIHNTHLFCVHTVPHDOLFRNPHQALHTANRPEDECVGEGGLACHOLCARGHCWPGPT 532
Db 467 LIHNTHLFCVHTVPMQDLFRNPHQALHTANRPEDECVGEGGLACHOLCARGHCWPGPT 526
Qy 533 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 592
Db 527 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 586
Qy 593 CAHYKDPFCVACRPSGVKPDLSYMBIWKFPDEEGACQPCPNCTHSCVDLDDKGPASQ 652
Db 587 CAHYKDPFCVACRPSGVKPDLSYMBIWKFPDEEGACQPCPNCTHSCVDLDDKGPASQ 646
Qy 653 RASPLTSIVSAVVGILLVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAM 712
Db 647 RASPLTSIVSAVVGILLVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAM 706
Qy 713 PNOAQMRILKETELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 772
Db 707 PNOAQMRILKETELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 766
Qy 773 ILDEAYVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCGLLDHVRENRRGLSGQDLLNWC 832
Db 767 ILDEAYVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCGLLDHVRENRRGLSGQDLLNWC 826
Qy 833 MOIAGMSYLEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGKVP 892
Db 827 MOIAGMSYLEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGKVP 886
Qy 893 KWALESILRRRPTHOSDVMSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQPI 952
Db 887 KWALESILRRRPTHOSDVMSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQPI 946
Qy 953 CTIDVYIMVWKMIIDSECRPRRELVSFERNARDPQRFVJQNEDELGPASPLDSTFYR 1012
Db 947 CTIDVYIMVWKMIIDSECRPRRELVSFERNARDPQRFVJQNEDELGPASPLDSTFYR 1006
Qy 1013 SLEDDMDGDLVDAEYLVPOQGFPCDPAPAGGMVHRHSSSTRSGGDLTLGLEPS 1072
Db 1007 SLEDDMDGDLVDAEYLVPOQGFPCDPAPAGGMVHRHSSSTRSGGDLTLGLEPS 1066
Qy 1073 EEEAPRSLAPSEGAGSDVFDGLGMAAKLQSLPHTDPSPLOQRYSEDTVPLPSETDG 1132
Db 1067 EEEAPRSLAPSEGAGSDVFDGLGMAAKLQSLPHTDPSPLOQRYSEDTVPLPSETDG 1126
Qy 1133 YVAPLTCSQPEYVNPQDVRPQPPSPREGPLPAARAGATLERPKTSLPGKNGVVKDVFA 1186
Db 1127 YVAPLTCSQPEYVNPQDVRPQPPSPREGPLPAARAGATLERPKTSLPGKNGVVKDVFA 1186
Qy 1193 FCGAVENBEYLTPOCGAARPOPHPPAFSAFNLYYMDODPPERGAPPSTFKGTPTAENP 1252
Db 1187 FCGAVENBEYLTPOCGAARPOPHPPAFSAFNLYYMDODPPERGAPPSTFKGTPTAENP 1246
Qy 1253 EYLGLDVPV 1261
Db 1247 EYLGLDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX AAR39568;
AC AAR39568;
XX XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
PN WO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
DR WPI: 1993-272889/34.
DR N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour antigen - for imaging or treating breast or ovarian cancer etc.
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of tumour cells, such as breast and ovarian tumour cells, which is an approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 97.3%; Score 6641; DB 14; Length 1433;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1232; Conservative 4; Mismatches 11; Indels 22; Gaps 2;

Qy 1 MELAALCRWGLLALLPPGAASQTCTGDMKRLPASBETHLDMLRHLYQGCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQTCTGDMKRLPASBETHLDMLRHLYQGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDQIEVQGVLIHQNVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDQIEVQGVLIHQNVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGASPGSLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGASPGSLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKGSRGCGESSEDQSLTRTVCAGGCARCKGPLPFNNFTVSF 240
Db 181 LTLIDTNRSRACHPCSPMCKGSRGCGESSEDQSLTRTVCAGGCARCKGPLP----- 232

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-250-264-14
Perfect score: 6839
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLCGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:*

1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|---------------------|
| 1 | 6702 | 98.0 | 1255 | 1 A24571 | protein-tyrosine k |
| 2 | 5895 | 86.2 | 1260 | 1 TVRTNU | protein-tyrosine k |
| 3 | 5880.5 | 86.0 | 1254 | 2 I48161 | p-185 precursor |
| 4 | 3116 | 45.6 | 1210 | 1 GQHUE | epidermal growth f |
| 5 | 3092 | 45.2 | 1210 | 2 A53183 | epidermal growth f |
| 6 | 3071.5 | 44.9 | 1223 | 1 TVCHLV | epidermal growth f |
| 7 | 2971.5 | 43.4 | 1308 | 2 A47253 | epidermal growth f |
| 8 | 2674 | 39.1 | 1166 | 1 S06142 | protein-tyrosine k |
| 9 | 2391.5 | 35.0 | 1342 | 2 A36223 | kinase-related tra |
| 10 | 2306.5 | 33.7 | 1339 | 2 JC4387 | epidermal growth f |
| 11 | 1766.5 | 25.8 | 698 | 1 TVFVLV | protein-tyrosine k |
| 12 | 1703 | 24.9 | 604 | 1 TVYUH | protein-tyrosine k |
| 13 | 1647 | 24.1 | 544 | 2 S35745 | protein-tyrosine k |
| 14 | 1640 | 24.0 | 545 | 2 S00727 | protein-tyrosine k |
| 15 | 1623 | 23.7 | 540 | 2 B44776 | kinase-related tra |
| 16 | 1621 | 23.7 | 540 | 1 TVFVEB | protein-tyrosine k |
| 17 | 1600.5 | 23.4 | 1330 | 1 GQFFE | epidermal growth f |
| 18 | 1484 | 21.7 | 644 | 2 A36325 | epidermal growth f |
| 19 | 1265 | 18.5 | 1323 | 2 S88257 | protein let-23 (im |
| 20 | 1265 | 18.5 | 1374 | 2 S70712 | protein-tyrosine k |
| 21 | 1171 | 17.1 | 1369 | 2 S70713 | protein-tyrosine k |
| 22 | 1146 | 16.8 | 1717 | 1 A45558 | epidermal growth f |
| 23 | 1103 | 16.1 | 527 | 2 A42032 | epidermal growth f |
| 24 | 945.5 | 13.8 | 843 | 2 A27131 | epidermal growth f |
| 25 | 866.5 | 11.8 | 346 | 2 S13807 | protein-tyrosine k |
| 26 | 754.5 | 11.0 | 311 | 2 S13808 | protein-tyrosine k |
| 27 | 733 | 10.7 | 1363 | 2 T43220 | insulin-like growth |
| 28 | 721.5 | 10.5 | 1372 | 2 A34157 | insulin receptor p |
| 29 | 719 | 10.5 | 1383 | 2 A36080 | insulin receptor p |

30 718 10.5 1382 1 INHUR insulin receptor p
31 691 10.1 1607 2 T43212 insulin-like growth
32 687 10.0 1477 2 T18534 protein-tyrosine k
33 685.5 10.0 1300 2 A36502 insulin receptor-r
34 675 9.9 1268 2 B36502 insulin receptor-r
35 649 9.5 1367 1 IGHURI insulin-like growth
36 635 9.3 1371 2 A33837 insulin-like growth
37 620 9.1 1390 2 T30346 insulin receptor -
38 611.5 8.9 2148 1 A56081 insulin receptor -
39 606 8.9 2101 2 S57245 insulin receptor (
40 594.5 8.7 987 2 A54092 protein-tyrosine k
41 589.5 8.6 977 2 S49004 tyrosine kinase Mp
42 588 8.6 1114 1 S05582 protein-tyrosine k
43 586 8.6 1091 2 S33596 protein-tyrosine k
44 584.5 8.5 976 2 A36355 protein-tyrosine k
45 583.5 8.5 952 2 I50612 protein-tyrosine k

ALIGNMENTS

RESULT 1

A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N:Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming protein e
C:Species: Homo sapiens (man)
C:Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.
Nature 319, 230-234, 1986
A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A:Reference number: A24571; MUID:86118663; PMID:3003577
A:Accession: A24571
A:Molecule type: mRNA
A:Residues: 1-1255 <YAM>
A:Cross-references: GB:X03363; MID:g31197; PIDN:CAA27060.1; PID:g31198
R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid
A:Reference number: A25491; MUID:86016729; PMID:2995967
A:Accession: A25491
A:Molecule type: DNA
A:Residues: 737-1031 <SEM>
A:Cross-references: GB:M11767; MID:g182163; PIDN:AAA35808.1; PID:g553282
R:Coussens, L.; Yang-Peng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg
Science 230, 1132-1139, 1985
A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro
A:Reference number: A44188; MUID:86070181; PMID:2999974
A:Accession: A44188
A:Molecule type: DNA
A:Residues: 740-910 <COU1>
A:Cross-references: GB:M12036; MID:g183988; PIDN:AAA35978.1; PID:g183989
A:Accession: B44188
A:Molecule type: mRNA
A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A:Cross-references: GB:M11730; MID:g183986
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A:Reference number: I59509; MUID:85272597; PMID:2992089
A:Accession: I59509
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 832-909 <REX>
A:Cross-references: GB:L29395; MID:g459807; PIDN:AAA35809.1; PID:g459808
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio
A:Reference number: I57622; MUID:87286898; PMID:3039351
A:Accession: I57622
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-191 <TAL>

A:Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C:Genetics:
A:Gene: GDB:ERBB2; NGL; NEU; HER-2
A:Cross-references: GDB:120613; OMIM:164870
A:Map position: 17q21.1-17q21.1
A:Introns: 25/1; 75/3; 147/1; 883/3
A>Note: the list of introns is incomplete
C:Function:
A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase

F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F:22-653/Domain: extracellular #status predicted <EXT>
F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
F:68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:886/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.0%; Score 6702; DB 1; Length 1255;
Best Local Similarity 98.3%; Pred. No. 8.3e-268; Indels 6; Gaps 1;
Matches 1240; Conservative 3; Mismatches 12;
QY 1 MELAALCRNGLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
DB 1 MELAALCRNGLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIEQVGVVLIHQNQVQVPLQRLIRVGTQLFEDNYALVDNG 120
DB 61 ELTYLPTNASLSFLQDIEQVGVVLIHQNQVQVPLQRLIRVGTQLFEDNYALVDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFPKNQLA 180
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFPKNQLA 180
QY 181 LTLIDNRRACHPCSPMKGSCWGESSEDCOSLTRTVCAGGCARCKGFLPTDCCHEQC 240
DB 181 LTLIDNRRACHPCSPMKGSCWGESSEDCOSLTRTVCAGGCARCKGFLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVFNFTVFWLRVPKVSASHLRYTFGAS 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVFNFTVFWLRVPKVSASHLRYTFGAS 300
QY 301 CVTACPNYVLTSDVGSCTLVCPHNOBVTAEQDQTCCKSKPCARVCYGLGMEHLREVR 360
DB 295 CVTACPNYVLTSDVGSCTLVCPHNOBVTAEQDQTCCKSKPCARVCYGLGMEHLREVR 354
QY 361 AVTSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQVQVETLEEITGYLYISA 420
DB 355 AVTSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQVQVETLEEITGYLYISA 414
QY 421 WDSLPLDSLVFQNLQVIRGILHNGAYSILTLQGLISWGLRLSLRGLSGLALIHNTHL 480
DB 415 WDSLPLDSLVFQNLQVIRGILHNGAYSILTLQGLISWGLRLSLRGLSGLALIHNTHL 474
QY 481 CFVHTVPDQLFRNPQALLHTANRPDECVGSLACHOLCARGHCWGPCPTQVCNCSOF 540
DB 475 CFVHTVPDQLFRNPQALLHTANRPDECVGSLACHOLCARGHCWGPCPTQVCNCSOF 534
QY 541 LRQGECEBCEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDP 600
DB 535 LRQGECEBCEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDP 594
QY 601 FCVARCPGVKPDLSYMPYWKPFDEEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSI 660

DB 595 FCVARCPGVKPDLSYMPYWKPFDEEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSI 654
QY 661 VSNVGVILLVNVLVGVVFGILLIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOQMRI 720
DB 655 ISAVGVILLVNVLVGVVFGILLIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOQMRI 714
QY 721 LKTELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVM 780
DB 715 LKTELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVM 774
QY 781 AGVSPYVRLIGLICLTSTVQLVTQMPYGCCLLDHVRENRLGSGDOLLNWCWOIAKMS 840
DB 775 AGVSPYVRLIGLICLTSTVQLVTQMPYGCCLLDHVRENRLGSGDOLLNWCWOIAKMS 834
QY 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESI 900
DB 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESI 894
QY 901 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICTIDVYMI 960
DB 895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICTIDVYMI 954
QY 961 MYKWMIDSECRPRPRELVSEFSRMARDPQRFVWIONEDLGPASPLDSTFYRSLLEDDDM 1020
DB 955 MYKWMIDSECRPRPRELVSEFSRMARDPQRFVWIONEDLGPASPLDSTFYRSLLEDDDM 1014
QY 1021 GDLVDAEYLVPOQGFCCPDPAAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1080
DB 1015 GDLVDAEYLVPOQGFCCPDPAAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1074
QY 1081 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDSPLQRYSEDPTVPLPSETDGVVAPLTC 1140
DB 1075 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDSPLQRYSEDPTVPLPSETDGVVAPLTC 1134
QY 1141 POPEYVNPQDVPRPQSPREGPLPAARPAAGATLERAKTISPGKQGVVQVFAFGGAVENP 1200
DB 1135 POPEYVNPQDVPRPQSPREGPLPAARPAAGATLERAKTISPGKQGVVQVFAFGGAVENP 1194
QY 1201 EYLTPOGGAAPOPHPPPAFSPADNLYYWDQPPPERGAPSTFKGPTAENPEYLGUDVP 1260
DB 1195 EYLTPOGGAAPOPHPPPAFSPADNLYYWDQPPPERGAPSTFKGPTAENPEYLGUDVP 1254
QY 1261 V 1261
DB 1255 V 1255

RESULT 2

TVRTNU

protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999

C:Accession: A24562; A61204

R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein

A:Reference number: A24562; MUID:86118662; PMID:3945311

A:Accession: A24562

A:Molecule type: mRNA

A:Residues: 1-1260 <BAR>

A:Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746

R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohe

A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals n

2-chiazolyl]formamide or N-methyl-N-nitrosourea.

A:Reference number: A61204; MUID:92035293; PMID:1682063

A:Accession: A61204

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 637-663, 'V', 665-702 <MAS>

A>Note: authors translated the codon GCA for residue 25 as Val

C:Genetics:

A:Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F;1-19/Domain: signal sequence #status predicted <SIG>
 F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
 F;658-680/Domain: transmembrane #status predicted <TMN>
 F;723-988/Domain: protein kinase homology <KIN>
 F;731-739/Region: protein kinase ATP-binding motif
 F;731.191.263.535.576.634/Binding site: carboxylate (Asn) (covalent) #status predicted
 F;691/Binding site: phosphate (Thr) (covalent) #status predicted
 F;758/Active site: Lys #status predicted
 F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.2%; Score 5895; DB 1; Length 1260;
 Best Local Similarity 86.4%; Pred. No. 1e-234;
 Matches 1091; Conservative 52; Mismatches 112; Indels 8; Gaps 3;

Qy 1 MELAALCRWGLLLALLPFGAASVQCTGTDMLRLPASPTHLDMLRLHYQCCVQGNL 60
 Db 4 MELAALCRWGLLLALLPFGIAGTQCTGTDMLRLPASPTHLDMLRLHYQCCVQGNL 63

Qy 61 ELYLPTNLSFLQDIEQVGYVLIANQVQVPLQRLIRVGTQLFEDNYALAVLNG 120
 Db 64 ELYVVPANASLFLQDIEQVGYVLIANQVQVPLQRLIRVGTQLFEDKYALAVLNR 123

Qy 121 DPLNNTTPTVT-GASPGGLRELQLSLTEILKGGVLIQRNPOL-CYODTILKWDIFHKNQL 179
 Db 124 DPQNVAASTGRTPEGLRELQLSLTEILKGGVLIQRNPOLCYQDMVLWKDVFRRKNQL 183

Qy 180 ALTILIDNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCGAGCARCKGPLPTDCCHEQ 239
 Db 184 APVDIDNRSRACHPCAPACKDNHCWGESPEDCQILTGTICTSGCCARCKGRLPTDCCHEQ 243

Qy 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVFNFTVFWLRVPKVSASHLRYTGA 299
 Db 244 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTNTDTFESMNP-----GRYTFGA 297

Qy 300 SCVTACPNYLSTDVSGCTLCPLHNOBVTAEADGTQRCCKSPKPCARCYGLGHEHREV 359
 Db 298 SCVTTCPNYLSTEVSGCTLCVPPNQBVTAEADGTQRCCKSPKPCARCYGLGHEHRLGA 357

Qy 360 RAVTSANIQEPAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFTLEITGYLYIS 419
 Db 358 RAITSNDNVQEPDGCKKIFGSLAFIPESFDGDPSSGIAPLRPEQLQVFTLEITGYLYIS 417

Qy 420 AWPDSLPLDSVFNQVIRGILHNGAYSLTQLGTSWGLRSLRELGSGLAIHNNHT 479
 Db 418 AWPDSLRLDSVFNQVIRGILHNGAYSLTQLGTHSLGLRSLRELGSGLAIHNNHT 477

Qy 480 LCFVHTVPWQDLFRNPHQALLHTANRPEDE-CVGEGLACHOLCARGHCWGPPTQCVNCS 538
 Db 478 LCFVHTVPWQDLFRNPHQALLHSGNRPEEDLCVSSGLVCNLSLCAHGCWGPPTQCVNCS 537

Qy 539 QFLRGQECVBCRVLQGLPREYVNHARHCLPCHPECPQNGSVTCFGEADQCACAHYKD 598
 Db 538 HFLRGQECVBCRVKGLPREYVSDKRLCPCHPECPQNSSETCFGEADQCACAHYKD 597

Qy 599 PPECVAPCPGVKPDLSYMPKWPDEBACQPCPINCSTHSCVDLDKGCAPRQASPLT 658
 Db 598 SSSCVAPCPGVKPDLSYMPKWPDEBACQPCPINCSTHSCVDLDKGCAPRQASPLT 657

Qy 659 SIYSVAVGILLVVLGVVFGILIKRQKIRKYMTRRLQETELVEPLTPSGAMPNQAQM 718
 Db 658 FIATVGVLLFLVLVVGILIKRQKIRKYMTRRLQETELVEPLTPSGAMPNQAQM 717

Qy 719 RILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAY 778
 Db 718 RILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAY 777

Qy 779 VMAGVGSPPYVSRLLIGICLTSTVQLVTOIMPYGCLLDHVRNRLGSGQDLLNWCQMIAKG 838
 Db 778 VMAGVGSPPYVSRLLIGICLTSTVQLVTOIMPYGCLLDHVRNRLGSGQDLLNWCQMIAKG 837

Qy 839 MSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP1KMWALE 898

Db 838 MSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP1KMWALE 897

Qy 899 SILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVY 958
 Db 898 SILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVY 957

Qy 959 MIMVKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPAASPLDSTFYRSLLEDD 1018
 Db 958 MIMVKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPAASPLDSTFYRSLLEDD 1017

Qy 1019 DMGDLVDAEYLVPOOGFFCDPPAPACAGMWHRRSSSTRSGGDLTLGLPSESEAPR 1078
 Db 1018 DMGDLVDAEYLVPOOGFFCDPPAPACAGMWHRRSSSTRSGGDLTLGLPSESEAPR 1077

Qy 1079 SPLAPSEGAGSDVFDGDLGMAAGLQSLPLTHDPSPLORYSDPTVPLPSETDGYVAPLT 1138
 Db 1078 SPLAPSEGAGSDVFDGDLGMAAGLQSLPLTHDPSPLORYSDPTVPLPSETDGYVAPLA 1137

Qy 1139 CSPQPEYVQPDVRPQPPSPREGPLUPAARAGATLERAKTLSPGKNGVVKDVAFAGGAVE 1198
 Db 1138 CSPQPEYVQSEVQPPPLTPPEGLPPVVRPAGATLERPKTLSPGKNGVVKDVAFAGGAVE 1197

Qy 1199 NPEYLTPOGGAAPQHPHPPAPSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLD 1258
 Db 1198 NPEYLVPREGTASPPHPSPAFDPNLYYWDQNSSEQPPSPNFEGTPTAENPEYLGLD 1257

Qy 1259 VPV 1261
 Db 1258 VPV 1260

RESULT 3
 I48161
 P-185 precursor - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C:Accession: I48161
 R:Nakamura, T.; Uehijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Iehika
 Gene 140, 251-255, 1994
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A:Reference number: I48161; MUID:94193007; PMID:7908275
 A:Accession: I48161
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1254 <RES>
 A:Cross-references: GB:D16295; NID:G493236; PIDN:BAA03801.1; PID:G747595
 C:Genetics:
 A:Gene: neu
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP
 F:718-983/Domain: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 86.0%; Score 5880.5; DB 2; Length 1254;
 Best Local Similarity 86.1%; Pred. No. 4e-234;
 Matches 1086; Conservative 60; Mismatches 108; Indels 7; Gaps 2;

Qy 1 MELAALCRWGLLLALLPFGAASVQCTGTDMLRLPASPTHLDMLRLHYQCCVQGNL 60
 Db 1 MELAALCRWGLLLALLPFGASGTVCTGTDMLRLPASPTHLDMLRLHYQCCVQGNL 60

Qy 61 ELYLPTNLSFLQDIEQVGYVLIANQVQVPLQRLIRVGTQLFEDNYALAVLNG 120
 Db 61 ELYLVPANATLSFLQDIEQVGYVLIANQVQVPLQRLIRVGTQLFEDKYALAVLNR 120

Qy 121 DPLNNTTPTVTGASPGGLRELQLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLA 180
 Db 121 DPLDNVTATGRTPEGLRELQLSLTEILKGGVLIQRNPOLCYQDTVLWKDVFRRKNQLA 180

Qy 181 LTILIDNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCGAGCARCKGPLPTDCCHEQC 240
 Db 181 PVDIDNRSRACHPCAPACKDNHCWGESPEDCQTLTGTIAPRAVPAARARLPTDCCHEQC 240

| | | | |
|----|------|---|------|
| Qy | 241 | AAGCTGPKGSDCLACLFHNSHGICELHCPALVFNNFTVSFWLVRVPKVSASHLERYTFGAS | 300 |
| Db | 241 | | |
| Db | 241 | AAGCTGPKGSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPE | 294 |
| Qy | 301 | CVTACPNYLSTDVGSCTLCPLHNOQVETADGTORCEKSKPCARVCYGLGMEHLREVR | 360 |
| Db | 295 | | |
| Db | 295 | CVTTCPNYLSTEVSCTLCPLNNOQVETADGTORCEKSKPCARVCYGLGMEHLRGAR | 354 |
| Qy | 361 | AVTSANIOEFAGCKKIFGSLAFLPDESFDGDPASNTAPLOPEQLOVFTLLEEITGLYLISA | 420 |
| Db | 355 | | |
| Db | 355 | AITSANIOEFAGCKKIFGSLAFLPDESFDGPNSSGIAPLTPLOEQLOVFTLLEEITGLYLISA | 414 |
| Qy | 421 | WPOSLLPDLVSFQNLQVLRGRIHLNGAYSLTLQGLGISWLGRLSRLRELGSGGLAIHHNTHL | 480 |
| Db | 415 | | |
| Db | 415 | WPOSLLHSLVSFQNLURVLRGRVLHGDAYSALQGLGIRWLGRSLURELGSLVLIRHNTHL | 474 |
| Qy | 481 | CFVHTVPWDQLFRNPHQALLHTANRPBDECVGEGLACHQLCARGHCGWGPPTQCVNCSQF | 540 |
| Db | 475 | | |
| Db | 475 | CFVHTVPWDQLFRNPHQALLHSGNPSBEECGKDXFACYPLCAHGHCGWGPPTQCVNCSHF | 534 |
| Qy | 541 | LRGQECVECBVLQGLPREYVNAHCLPCHPECOPQNGSVTCRGPPEADQCVACAHYKDPP | 600 |
| Db | 535 | | |
| Db | 535 | LRGQECVKECRVWKGKLPREYVNGKHCLPCHPECOPQNSTETCTGTSEADQCTACPHYKDSP | 594 |
| Qy | 601 | FCVARCPSGVKPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKCCPAEQRASPLTSI | 660 |
| Db | 595 | FCVARCPSGVKPDLSYMPIWKYPDEEGWCOPCPINCTHSCVDLDERCCPAEQRASPATSI | 654 |
| Qy | 661 | VSAVVGILLVVVLGVGFGLIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQOMRI | 720 |
| Db | 655 | | |
| Db | 655 | IATVVGILLFLVIGVVGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQOMRI | 714 |
| Qy | 721 | LKETELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM | 780 |
| Db | 715 | | |
| Db | 715 | LKETELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM | 774 |
| Qy | 781 | AGVSPVYSRLGLICLTSVTQVLTPQMPYGCCLDHDVRENRGRLGSDOLLNWCQIAKMS | 840 |
| Db | 775 | AGLSPVYSRLGLICLTSVTQVLTPQMPYGCCLDHDVREHGRGLGSDOLLNWCQIAKMS | 834 |
| Qy | 841 | YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALES | 900 |
| Db | 835 | YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWIALES | 894 |
| Qy | 901 | LRRRFTHQSDVWSYGVTVWELMTFGAKYPDGIIPAREIPDLLEKGERLPQPICTIDVYMI | 960 |
| Db | 895 | LRRRFTHQSDVWSYGVTVWELMTFGAKYPDGIIPAREIPDLLEKGERLPQPICTIDVYMI | 954 |
| Qy | 961 | MVKCWMIDSECRPRFRELVSSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM | 1020 |
| Db | 955 | MVKCWMIDSECRPRFRELVSSEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDDDM | 1014 |
| Qy | 1021 | GDLVDAEYILVPOQGFPCPDAPCAGCMVHRRHSSSTRSGGDLTLGLEPSEEEAPRSP | 1080 |
| Db | 1015 | GDLVDAEYILVPOQGFPPDAPCAGTAHRRHSSSTRSGGELTLGMEPSEEPSPRSP | 1074 |
| Qy | 1081 | LAPSEGAGSDVFDGLGMGAAGKLSLPTHDPSPQLQRYSEDPTVPLPSETDGVVAPLTC | 1140 |
| Db | 1075 | LAPSEGAGSDVFEGELCMGATKGQSI:SRDLSPLQRYSEDPTLPLTPETDGVVAPLACS | 1134 |
| Qy | 1141 | POPEYVNOQDVRPOPPSPRGGPLPAARPAGATLERAKTLSPGKNGVVKQVFAFGGAVENP | 1200 |
| Db | 1135 | POPEYVNOQEVPRPOPPPLTPGGPLPPVPRPAGATLERPKTLSPGKNGVVKQVFTFCGAVENP | 1194 |
| Qy | 1201 | EYLTPOCGAAAPQHPHPPAFPAFDNLVYWDQDPPERCAPPSTFKGTPAENPYVLGLDVP | 1260 |
| Db | 1195 | EYLVPRGGSASQPH-PPALCPAFDNLVYWDQDPSERGPSNTFEGTPAENPYVLGLDVP | 1253 |
| Qy | 1261 | V 1261 | |
| Db | 1254 | ↓ 1254 | |

RESULT 4

QOHUE

N:Contains: epidermal growth factor receptor precursor - human
C:Species: Homo sapiens (man)
C:Date: 15-Nov-1994 #sequence revision 27-Nov-1985 #text change 11-Jun-1999
C:Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143;
R:Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tash, A.W.; Lee, J.;
rg, P.H.
Nature 309, 418-425, 1984
A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression
A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <LL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A:Note: the authors translated the codon AAG for residue 540 as Asn
R:Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <SH>
A:Cross-references: GB:M1234; NID:g181981; PIDN:AAA23370.1; PID:g553272
R:Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termi
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A:Experimental source: carcinoma cell line A431-7
R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.;
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RN
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 150-187, 'KSVIOAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-
'798-799', 'TD', 802-811, 'R', 813-942 <XUY>
R:Lin, C.R.; Chen, W.S.; Kruijer, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma,
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplificati
A:Reference number: A43615; MUID:84196372; PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.I.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172103; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>

R; Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
A:Reference number: A60143; MUID:85182650; PMID:2985580
A:Accession: A60143
A:Molecule type: protein
A:Residues: 740-744, 'X', 746-747 <RUS>
R:Mróczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
A:Reference number: A38023; MUID:84191554; PMID:6325948
A:Contents: annotation; receptor activity
A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R:Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
Cell 59, 33-43, 1989
A:Title: Functional independence of the epidermal growth factor receptor from a domain Z
A:Reference number: A3331; MUID:90003233; PMID:2790960
A:Contents: annotation; internalization signal
C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor
C:Genetics:
A:Gene: GDB:EGFR
A:Cross-references: GDB:120610; OMIM:131550
A:Map position: 7p12.3-7p12.1
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-1210/Product: EGF receptor #status predicted <MAT>
F:25-645/Domain: extracellular #status predicted <EXT>
F:75-300/Domain: EGF receptor extracellular domain repeat <BE1>
F:390-600/Domain: EGF receptor extracellular domain repeat <BE2>
F:646-668/Domain: transmembrane #status predicted <TM>
F:669-1210/Domain: intracellular #status predicted <INT>
F:710-975/Domain: protein kinase homology <KIN>
F:718-725/Region: protein kinase ATP-binding motif
F:999-1046/Region: coated-pit mediated internalization signal
F:1047-1210/Region: inhibitory
F:128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic
F:745/Active site: Lys #status experimental

Query Match 45.6%; Score 3116; DB 1; Length 1210;
Best Local Similarity 49.2%; Score No. 9e-121;
Matches 625; Conservative 179; Mismatches 355; Indels 112; Gaps 22;

Qy 11 LLLALLPPGAA--STQVCTGTDMLRLPASPETHDMLRHLQGCQVQVGNLEITYLPTN 68
Db 14 LLAALCPASRALEKKVKCGTSNKLTLQGTGFEDHFLSLQRMFNCEVVLGNLEITYVQRN 73
Qy 69 ASLSFLDIOIEVOGVYLIHNOVRVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNNTTP 128
Db 74 YDLSFLKTIQEVAGVYLIHNTVERIPLENLQIIRGNMYYSYALAVLSNYD----- 126
Qy 129 VTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNQALALTIDTNR 188
Db 127 ---ANKTGLKELPMNLQELILHGAFRFSNNPALCNVESIQWRDIVSSDFLSNMGMDFQNH 183
Qy 189 SRACHPCSPMKGSRGWGSESDCQSLTRTVCAAGCA-RCKGPLPTDCCHCEQCAAGCTGP 247
Db 184 LGSCQKCDPSPNGSCWGAAGEENCQKLTIIICAQCSGRCGRKSPDCCHNQCAAGCTGP 243
Qy 248 KHSCLACLHNHSGICELHCPALVFNNTVSVFLRPVKVSASHLRYTFGASCVTACPY 307
Db 244 RESDCLVCRKTRDEATCKDTCPLMLNPTTY-----QMDVNPGEKYSFGATCVKCKPR 297
Qy 308 NYLSTDVGSCTLVCPHLNHOEVAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSANI 367
Db 298 NYVVTDHGSCVRACGADSVEM--EEDGVKCKKEGCPKVCNGIGIGEFKDSLSINATNI 356
Qy 368 QEPAGCKKIFGSLAPLPSFGDGPASNTAPLQPEQLQVFTLEETGYLYISAWPDSLDP 427
Db 357 KHFNKCTSIISGDLHTLPVAFRGDSFTHTPPLDPQELDKTKVKEITGTFLLIQAWPENRTD 416
Qy 428 LSVFONLQVIRGILHNGAYSILTLOGLGTSWLGLRSIRLGLSGLALIHNTLHLCFVHTVP 487
Db 427 LSVFONLQVIRGILHNGAYSILTLOGLGTSWLGLRSIRLGLSGLALIHNTLHLCFVHTVP 487

RESULT 5

RESULT
A53783

epidermal growth factor receptor precursor - mouse

C:Species: *Mus musculus* (house mouse)

```
C:\Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
```

C;Accession: A53183; A43818; S24942; A28941; S45325: I49643

R;Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.

Genes Dev. 8, 399-413, 1994

A;Title: The mouse waved-2 phenotype results from a point mutation in the *Wnt3* gene.

A;Reference number: A53183; MUID:94170986; PMID:8125255

A;Accession: A53183

A;Molecule type: mRNA

A;Residues: 1-1210 <LUE>

A;Cross-references: GB:U03425

R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
A:Reference number: A43818; MUID:91232866; PMID:2030916
A:Accession: A43818
A:Molecule type: mRNA
A:Residues: 1-714 <AVI>
A:Cross-references: GB:X59698
R:Eisinger, D.P.; Sertero, G.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24942
A:Accession: S24942
A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <BIS>
A:Cross-references: EMBL:Z12608
R:Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A:Reference number: A28941; MUID:88330814; PMID:3138233
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000; 1002-1009,
R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: S45325
A:Accession: S45325
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse B
A:Reference number: I49643; MUID:93126380; PMID:7678348
A:Accession: I49643
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 22-132 <RES>
A:Cross-references: GB:L06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
C:Genetics:
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphopro
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:728-728/Domain: protein kinase ATP-binding motif
F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.2%; Score 3092; DB 2; Length 1210;
Best Local Similarity 49.1%; Pred. No. 8.7e-120;
Matches 628; Conservative 171; Mismatches 363; Indels 116; Gaps 24;

Qy 11 LLLALLPGAA--STQVCTGDKMLRLPASBETHLDMLRLHYQGVQVQGNHLELYLPTN 68
Db 14 LTLTCAAGGAALEKKVCGQSNRUTQLGTGFEDHFLSLQRMVNNCEVVLGNLEITYQVN 73
Qy 69 ASLGFQDIQEVQGVVLIHNOVQLQRLIRVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db 74 YDLFLTKIQEVAGVLIATNTVERIPLENQIRGNALYENTYALAILSN----- 124
Qy 129 VTGASPGGLRELQRLSLTEILKGVLIQORNPOLCYQDTILMKDI----FHKNNQALTLI 184
Db 125 -YGNRTGLRELPMRLNQLLEILLIGAVRFNSNNPILCNMDTIQWRDIQVNVFMSNMSDL--- 180
Qy 185 PTNRSRACHPCSPMKCKSRGCESEDCSLTRTVCAGGCA-RCKGPLPTDCCHQCAAG 243
Db 181 -QSHPSCPKCDPSPNGSCWGGGEENCOKTKIICAAQCSHRRCGRSPSCCHQCAAG 239

Qy 244 CTGPKHSDCLACLFHNSGICELHCPALVFNFTVSFMLRVPKVSASHLRYTFGASCVT 303
Db 240 CTGPRESCLVCQKQFQDEATCKDTCPPLMLNPTTY-----QMDVNEPKGKSFQATCVK 293
Qy 304 ACPNYLSTDVSGTCLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAT 363
Db 294 KCPRNVVYTDHSGCVACGPDYEV-EEDGIRKCKKCDGCPCKVCNGIGIGISFKDITLSIN 352
Qy 364 SANTQEPAGCKKIFGSLAFLPESPDGDPASNTAPLOPEQLQVVFETLEITGYLYISAWPD 423
Db 353 ATNKHFKYCTAISGDHLIPVAFKGDSTFTPLDPLRELEILKTKVKEITGELLQAWPD 412
Qy 424 SLPDLVSFONLQVIRGRILHNGAYSLTLOGLGISWGLRSRLSELGSLALIHHTHLCFV 483
Db 413 NWTDLHAFENLEIRGRTKHQGFSLAVVGLNITSLGLRSLEIKESDGVIIISGRNRLCYA 472
Qy 484 HTVPMWQLFRPHQALLHTANRPEDECVGEGELACHQLCARGHCWGGPTQCVNCQSLR 543
Db 473 NTINWKLFGTPNQKTKIMNRAEKDKAVNHVCNPLCSSEGCWGPEDRCVSCQNVSRG 532
Qy 544 QECVEECRVLOGLPREYVNAHCLPCHPECOPOGSGVTCFGEADQCVACAHYKDPFCV 603
Db 533 RECVEKCNILEGEPREFVENSECQCHPECLPQAMNITCTGRGPONCICQAHYIDGPHCV 592
Qy 604 ARCPGKVPDLSPYMPKFPDEEGACQPCINCTHSCVDLDDKGCPEAQSRASPLTSIVA 663
Db 593 KTCFAGIMGENTL-VWKYADANNVCHLCHANCTYGCAGPLQGCVEWPSGPKISIAFG 651
Qy 664 VVGLLVVLGVVFGI-LIKRRQOKIRKYTMRRLLQETLEVEPLTPSGAMPNQAOIRLIK 722
Db 652 IVGGLLFIVV-VALGIGLFMRHRHVRKXTRLLQLERLEVEPLTPSGEAPNQAHILIK 710
Qy 723 ETELRKVKVLSGAGFTYKGIWIPDGNVKIPVAIKVLRNTPSKANKEILDEAYVMAG 782
Db 711 ETEFKKIKVLSGAGFTYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMAS 770
Qy 783 VGSPPVSRLLGILCTSTVOLVTLMPYGLLDHVRNRLGSDLLANCMQIAKGMVYL 842
Db 771 VDNHVCRLGLICTSTVOLITQLMPYGLLDYVREHKONIGSQVLLNWCVQIAKGMVYL 830
Qy 843 EDVRLVHRDLAARNVVKSPNHVKITDFGLARLLIDIDETEHADGCKVPKIMWALESLR 902
Db 831 EDRLVHRDLAARNVVKTPQHVKITDFGLAKLGAEEKEYHAEGKVPKIMWALESLILH 890
Qy 903 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIW 962
Db 891 RIYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASDISSILEKGERLPQPPICITIDVYMIW 950
Qy 963 KCMWIDSCRPRFRELVSERMDARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDDMG 1021
Db 951 KCMWIDASRPKRELILIEFQOMARDPQRYLVIQODERHMLPSPDTSNPFYRALMDEEDME 1010
Qy 1022 DLVDAEYLVPOQGFCCPDPAFGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAAPSPL 1081
Db 1011 DVVDADEVLIPOQGF-----NSPST-----SRTPL 1036
Qy 1082 APSEGAGSDVFDGLGMGAAGLQSLPHTHDSPLQRYSEDTVPVLPSET--DGYVAPLTC 1139
Db 1037 LSSLATSIN---NSTVACINRNGSCRVKEDAFQRYSSDPTGAVTEDIDDAFL----- 1087
Qy 1140 SPOEYVNOQDVRPOPPSPREGPLPAAPAGATLERAKTLPCKNGVVKDVPFAGFAGVEN 1199
Db 1088 -PVEYVNOQ-SVPKRPAGSVQNVTHNQPLHP-----APGRDLHYON---PHSNVGN 1135
Qy 1200 PEYL-TPQGAAPQHPHPPAFSPAFDNLVYWDQ-----DP-----PERCAPST 1242
Db 1136 PEYLNTAQ-----PTCLSSGFSNFPALIKGSHQMSLDNPYQDFFPKETPKNGI 1186
Qy 1243 FKGTPTAENPYGLDVP 1260
Db 1187 FKGTPTAENPYGLDVP 1203

A:Title: Ligand-specific activation of HER4/p180erbB4 a fourth member of the endomorphin receptor family. Recd: Oct. 6.0.0.0: 20, 1740-1750, 1993

A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PJO>
A:Cross-references: GB:L07868; NID:g337359; PIDN:AA859446.1; PID:g337360
A:Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Query Match 43.4%; Score 2971.5; DB 2; Length 1308;
Best Local Similarity 45.2%; Pred. No. 8.2e-115;
Matches 612; Conservative 187; Mismatches 373; Indels 181; Gaps 30;

QY 9 WGLLLALLPPGNA-----STQVCTGTDMLRLPASBETHLDMLRHLYOCQVQGNLEITY 64
DB 8 WWMVSLVAAGTVQPSOSVQAGTENKSLSSLDLEQQYRALRYKCYENCEVVMGNLEITS 67
QY 65 LPTNASLSFLODIOEQVGYLIAHQVQVLPQRLRIVRGTOLEFEDNYALAVLDNGDPLN 124
DB 68 IEHNRDLSFLSVREVTGYVLVALNQFYLPLENRIIRGTLYEDRYALALFLNRYKDG 127
QY 125 NTPPTVGASPGGLRELRLSLTEILKGGVLIQRPQLCYQDTILMKDIFHKNNQALTLI 184
DB 128 NF-----GLQELGLKNLLEILNGGVYVQDKFLCYADTIHQDIVRNWPNSNLTIV 178
QY 185 DTRSRACHPCSPMKGSRWCSESSEDQSLTRTVACGC-ARCGPLPTDCHCHQCAAG 243
DB 179 STNGSSGGRCHKSCGTG-RCMGPTENHCQTLTRTVCAEQCGRCGYGVYSDCHRECAAG 237
QY 244 CTGPKHSDCLACLHFNHSGICELHCP-ALVFNNFTVSFWLRVPKVSASHLRYTFGASCV 302
DB 238 CSGPKDTCFACMNDNSGACVTCQPTQFVYNTTF-----QLEHFNKAYITTGACV 290
QY 303 TACPYNILSTDVGSCTLVCPLNHQNVTADGTQRCCKSKPCARVCYGLGMEHLREVRV 362
DB 291 KCKPHNFV-VDSSSCVACRCPSSKMEV-BENGIMCKPCTDIPCACDGTGTSLSMAQTV 348
QY 363 TSANIOEAGCKKIGSLAFLPESFDGDPASNTAPLOEQLOVFTLEBITLYLISAWP 422
DB 349 DSSNIDKFINCTKNGNLIFLVTGTHGDPYNAIEAIDPEKLVNFTVREITGFLNIQSWP 408
QY 423 DSLPDSYFQNLQVIRGRIHNGAYSLTLQGLIGISWGLRSLRELGSGLALHHTHLCF 482
DB 409 PNMTDFVSFLVTGVRVLSGLSLLILKQGITSLQFSKLSAGNIYITDNSLCY 468
QY 483 VHTVPWDOLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWPGPTQCVCNCSQFLR 542
DB 469 YHTINWTLFTSTINQIRVIRDNKAENCTAEGMVCNHLCSDDCGWPGPDQCLSKRFSR 528
QY 543 GOECVEECRVLOGLPREYVNAHCLPCHPECOQ-ONGSVTCRPEADOCVACAHYKDPFP 601
DB 529 GRICIESNLVDFEFRENGSICVECDPQCEKMDGLTCHGPGPDNCTKCSHPKDFPN 588
QY 602 CVARCPGKVPDLSYMPYKFPDESGACOPCPINCTHSCVOLDKGC-----PA 650
DB 589 CVEKCPDGLQANSF--IFKYADPRECHPCNCTCGNGPTSHDCIYYPWTHGSHLTPQ 646
QY 651 EORASPLTSIVAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPS 709
DB 647 HAR-TPL--IAAGVTGGLFILVIGLTAFAVYVRKSIL-KKRALRRFL-ETELVEPLTPS 701
QY 710 GAMPNQAQMRILKETELRVKVLGSGAGTYVKGITWIPDGENVKIPVAIKVLRNTSPKA 769
DB 702 GTAPNQAQRLIKETELKRVKVLGSGAGTYVKGITWIPDGENVKIPVAIKVLRNTSPKA 761
QY 770 NKEILDEAVMAGVSPVSRLLGLCLTSTVOLVTQLMPYGCLLDHVRENRCRLGSDQL 829
DB 762 NVEFDEALIMASMDPHLVRLLGLVCLSTTQLVTLQMLPHGCLLEYVHEHKDNIQSULL 821
QY 830 NWCWOIAKGMVLEBRRVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGK 889

DB 822 NWCWOIAKGMVLEBRRVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGK 881
QY 890 VPIKWMALLESILRRRFTHQSDVWSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPO 949
DB 882 MPIKWMALLECITHYRKFTHQSDVWSYGVTTWELMTFGGPKYDGIPTREIPDLLEKGERLPO 941
QY 950 PPICTIDVYIMVVKCMIMIDSECRPRFRELVSFSESMARDPQRFVFIQNEED-LGPASPLDS 1008
DB 942 PPICTIDVYIMVVKCMIMIDADSRPKFELAAEFSESMARDPQRYLVLIQDDRMKLPSPNDS 1001
QY 1009 TFYRSLLEDDMGDLVDABEYLVPQGFPCPDPAFGAGMVMHRRSSSTRSGGDLTLTG 1068
DB 1002 KFFQNLLEDEEDLMDMAEYLVP-QAFNIPPP-----IYTSRAIDSNRS-----EIG 1049
QY 1069 LEPSEEEAPRS-----PLAP-SEGAGSDVFDGDLGNG 1099
DB 1050 HSPPPPAYTPMSGNOFYVYRDGGFAAGVSVVPYRAPSTIPEAPVAGATAEIFDDSCCNG 1109
QY 1100 AAKGLQSLPTHDPSPLOQYSEDPVPLPS-----ETDGYVAPLTCSPQPEVYNQPDVR 1152
DB 1110 TLRKPVAVHVOEDSTQRYSDPTVFAPERSPRGELDEEGYMTPRDKPKQEYLNVPV-- 1167
QY 1153 PQQPSREGPLPAARPAAGATLERAKTSLPGKNGVVKVFAFGAVENPEYLTPOGGAAPQ 1212
DB 1168 -----ENPFVSR--KNGDLQ-----ALDNPYHNASG---- 1194
QY 1213 PHPPPA-----FSPAFDNLVYDQDPPPERGA 1238
DB 1195 --PPKAEDEVYNEPLVNTFANTGLKAEYLNILSMPEKAKKAFDNDPDYMHSLPPRST 1252
QY 1239 --PRSTFGTPT-----AENPEYL 1255
DB 1253 LQHPDYLOEYTKYFKQNGRIRPIVAENPEYL 1285

RESULT 8

S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N:Alternate names: epidermal growth factor receptor homolog; kinase-related transform:
C:Species: Xiphophorus maculatus (southern platyfish)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C:Accession: S06142; S13809
R:Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Robe:
Nature 341, 415-421, 1989
A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu:
A:Reference number: S06142; MUID:90015140; PMID:2797166
A:Accession: S06142
A:Molecule type: DNA
A:Residues: 1-1166 <WIT>
A:Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R:Adam, D.; Maeueller, W.; Scharlt, M.
Oncogene 6, 73-80, 1991
A:Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphoph:
A:Reference number: S13807; MUID:91125882; PMID:1846957
A:Accession: S13809
A:Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 821-1025, N', 1027-1098, A', 1100-1166 <ADA>
A:Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C:Genetics:
A:Gene: mrk
A:Map position: Y
A:Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/3
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; t:
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif
Query Match 39.1%; Score 2674; DB 1; Length 1166;
Best Local Similarity 45.0%; Pred. No. 1.2e-102;


```

579  FRDPHCVNCSPHILG--AKGPIYKPYDAQNECRPCHENCQTQCNGPEIQDCLGQAEVL 636
654  ASPLTSIVSAVVGILLVVVLGWGFEIILIKRQOKIR-KYTMRRLLQETELVEPLTPSGAM 712
637  MSKPHLVIAVTVG--LAVILMLGGSFLYWRGRIQNKRAMRRYLERGESIEPLDPS-EK 693
713  PNOAQMRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 772
694  ANKYLARIKETEELURKULGVSGVFTGKHGIWIPGESIKIPVCIKVIEDKSGRQSFQA 753
773  ILDEAYVMAGVSGPSVSRLLGICLTSTVQLVTQLMPPYGCLLDHVRENRGRIGSDQLNWC 832
754  VTDHMLAVGSLDHAHIVRLGLGCPGSSILQVTVQVLPGLSLLDHVQKHRETLPQALLNWG 813
833  MQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPPI 892
814  VQIAKGMYLEEHSVMHRDLALRNVMLKSPSQVADFGVADLLPDPDKQLLHSEAKTPI 873
893  KWMALESI LRFRFTHQSDVMSYGVTVNMLMTFGAKPYDGPAREIPDLLEKGERLPPOPI 952
874  KWMALESIHFCKYTHQSDVMSYGVTVNMLMTFGAEPYAGLRAEIPDLLEKGERLAPOI 933
953  CTIDVYMWKCMWIDSECRPRFELVSEFSRMDARDPQRFVVIQNEDLGPASPLDSTFYR 1012
934  CTIDVYMWKCMWIDENIRPTFKELANEFTRMARDPPRLVVIKRAS-GPGTP--PAAEP 990
1013  SLLEDDMGDLVDAAEYVLVPOQGFCDPPAGCMGVHHRHSRSTSGGGDLTLGLGEP 1072
991  SVLTTKEL-----QEALEPEL-----DLDLDEAE 1016
1073  EE-----EAPRSP LAPSEG-----AGSDVFDGLDGMGAAGL 1104
1017  EELGATSLGALSLSLTGTLTRPGCSLLSPSSGYMPNMQSSLGAECLDSAVLGRGQFS 1076
1105  QSLPETHDPSPLQRYSEDPVLPSETQGV-----APL-----TC-----SPOPE---- 1144
1077  RPISLH-PIPRGR-----PASESSEGHVTGSEAELOEKVSVCKRSRSPRPRGDSA 1128
1145  VYNOQDVRPOPPSPREGF-----LPAARPAGATLERAKTLP-GKNGVV----- 1187
1129  YHSQRHSLTFVTPLSPGLLEEEDCGNGYVMPDTHLRGASSREGTLLSVGLSSVIGTTEE 1188
1188  -KOVFAFGGAVENPEYLTPOGGAAPQHPHP 1216
1189  DED-----EYEYMNKRKRGSP-PRPP 1209

```

RESULT 11

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and p
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NLI>
A;Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: In Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif

P:229/Active site: Lys #status predicted

| Query Match | 25.8% | Score 1766.5 | DB 1 | Length 698 | |
|-----------------------|-------|--|----------------|------------|---------|
| Best Local Similarity | 52.2% | Pred. No. 1.1e-65 | | | |
| Matches | 374 | Conservative 80 | Mismatches 137 | Indels 125 | Gaps 18 |
| Qy | 584 | GPBADOCVCAHYKDPFFCVARCPGVKPDLSYMPITWKPDEGACOPCPINCTHSCVDL | 643 | | |
| Db | 60 | GP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VMKYADANAVCOLCHPNCTRCKG | 116 | | |
| Qy | 644 | DDRGCPAEOPASPLTGISAVV-GILLVVLGVVFGILIKRROKIRKYTMRLLOSTEL | 702 | | |
| Db | 117 | GLEGCP--NGSKTPSIAAGVGGLLCLVVGLIGLYLRRR-HIVKRTLRLLOREL | 172 | | |
| Qy | 703 | VEPLTSGAMPNOAQMRLLKETELRKVKVLGSGAFYVYKGIWIPGENVKIPVAIKVL | 762 | | |
| Db | 173 | VEPLTSGEAPNOAHLRIILKETEFKKVKVLGSGAFYVYKGLWIPGEKVKIPVAIKEL | 232 | | |
| Qy | 763 | ENTSPKANKEILDEAYMAGVGSPPYRLLGLCTSTVOLVLTOLMPYGCLLDHVREN | 822 | | |
| Db | 233 | EATSPKANKEILDEAYMASVONPHVCRLLGLCTSTVOLITOLMPYGCLLDYIREH | 292 | | |
| Qy | 823 | LGSOQLLNWCQITAKGMSYLEVRLVHRDLAARNVLVKSPNHVKITDFGLARLLD | 882 | | |
| Db | 293 | IGSOQLLNWCQITAKGMVLEERLLVHRDLAARNVLVKTPOHVKITDFGLAKLLGA | 352 | | |
| Qy | 883 | YHADGKVPKIMWALSILRRRTHOSDVWSGYVTWMLTFGAKYDGIPIAREIPDL | 942 | | |
| Db | 353 | YHAEGGKVPKIMWALSILHRIYTHOSDVWSGYVTWMLTFGSKPYDGIPIASEIS | 412 | | |
| Qy | 943 | KGRLPQPPTCTIDVYIMVVKWCMIDSECRPRELVSFSRWARDPQRFVITQ-NED | 1001 | | |
| Db | 413 | KGRLPQPPTCTIDVYIMVVKWCMIDADSRKPRELIAEFSKWARDPPRYLVITQ | 472 | | |
| Qy | 1002 | PASPLDSTFYRSLLEDDMGDLVDAEYLVPOGGFFCPDAPGACGMVHRRHSSSTR | 1061 | | |
| Db | 473 | LPSPDTSKFYRLTMEEDMEDIVDAEYLVPHQGF-----NSPST--- | 513 | | |
| Qy | 1062 | GGDLTLGLEPSEEEAPRSP-----APSGAGSDVFDGDLGMAAKGLSLPHTD | 1116 | | |
| Db | 514 | -----SRTPLLSLSATSNSATNCID-----RNGQGHVPVRESFVQ | 550 | | |
| Qy | 1117 | RYSEDPTVPLPSET--DGVVAPLTCSPQPEYVNOPDVRPQPPSPREGPLPAAR | 1174 | | |
| Db | 551 | RYSDPTGHNLEESIDOGFL-----PAPEYVNQ--LMPKKPS----- | 595 | | |
| Qy | 1175 | RAKTLSPGKNGVVKYVF-----AFGAVENPEYLTPOGGAAPOPHPPPAF | 1219 | | |
| Db | 586 | ----TAMVQNIYNNISLTAISKLPWDSRYQNSHSTAVDNPEYL----- | 633 | | |
| Qy | 1220 | SPAFDNLVYWDQ-----DPPE-----RGAPPSTFKGTPTAENPEYGLDVP | 1260 | | |
| Db | 634 | KTVFESSPYIISGNNHQINLNDPQYQDFLPNETKPNGLLYKPAENPEYLRVAAP | 689 | | |

RESULT 12

TVYUH

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C/Species: avian erythroblastosis virus
C/Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C/Accession: A00644; A38022
R/Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A/Rittle: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A/Reference number: A00644; MUID:84026539; PMID:6313229
A/Accession: A00644
A/Molecule type: DNA
A/Residues: 1-604 <YAM>
A/Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R/Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin
Science 224, 1456-1459, 1984
A/Rittle: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type c
A/Reference number: A38022; MUID:84223957; PMID:6328658

A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W
A;Cross-references:
C;Genetics:

| | | | | |
|-----------------------|--------------|------------------|-----------------|-------------|
| Query Match | 24.9% | Score 1703; | DB 1; | Length 604; |
| Best Local Similarity | 52.2% | Pred. NO. 4e-63; | | |
| Matches 360; | Conservative | 76; | Mismatches 128; | |
| | | | Indels 126; | Gaps 16; |

RESULT 13
S35745

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C;Accession: S35745
R;Vennstroem, B.
submitted to the EMBL Data Library, March 1993

A;Reference number: S35743
A;Accession: S35745
A;Molecule type: DNA
A;Residues: 1-544 <VEN>
A;Cross-references: EMBL:X
C;Genetics:

| | | | | |
|-----------------------|--------------------|-------------|-----------------|-------------|
| Query Match | 24.1% | Score 1647; | DB 2; | Length 544; |
| Best Local Similarity | 54.9%; | | | |
| | Pred. No. 7.le-61; | | | |
| Matches 345; | Conservative | 70; | Mismatches 121; | Indels 92; |
| | | | | Gaps 15; |

RESULT 14
S00727

kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
R;Accession: S00727
R;Scotting, P.; Vennartom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mut;
A;Reference number: S00727; MUID:88217326; PMID:2897102
A;Accession: S00727

A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: Atp; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

| | | | | | | | | | |
|-----------------------|--------|--------------|----------|------------|--------|--------|-----|------|-----|
| Query Match | 24.0%; | Score | 1640; | DB 2; | Length | 545; | | | |
| Best Local Similarity | 54.9%; | Pred. No. | 1.4e-60; | | | | | | |
| Matches | 345; | Conservative | 69; | Mismatches | 122; | Indels | 92; | Gaps | 15; |

| | | | |
|----|------|--|------|
| Qy | 584 | GPEADQCVACAHYKDPFCVCARCPGKVPDLSYNPIWKFPDEGACQCPFCINCHTSCVDL | 643 |
| Db | 1 | GP--DHCMKCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCQLCHPNCRTGCKGP | 57 |
| Qy | 644 | DDKGCPAEQRASPLTSTVSAAV--GILLAVVLGVVFGILIKRRQOKIRKYTMRLLOSTEL | 702 |
| Db | 58 | GLGSCP---NGSKTPSIAAGVVGGLCLLVVGLIGIGLYLRR--HIVKRTLRLLQLQREL | 113 |
| Qy | 703 | VEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTGYKGIWIPDGENYKIPVAIKVL | 762 |
| Db | 114 | VEPLTPSGEAPNQAHLRLKETEFKKVKVLGFGAFGTVYKGLMPEGEKVTIPVAIKEL | 173 |
| Qy | 763 | ENTSPKANKELDEAYVMAGVGSPPYSRLLIGLICLTSTVQLVTLMPYGCLLDHVRENRR | 822 |
| Db | 174 | EATSPKANKELDEAYVMASVDNPHVCRLLIGLICLTSTVQLITQLMPYGCLLDYIREHKD | 233 |
| Qy | 823 | LGSQDLLNMCWQIAKGSYLEDLVRLVHRDLAARNVLKVSNNHVKITDFGLARLLDIDETE | 882 |
| Db | 234 | IGSQYLLNMCVQIAKGNVYLEERHLVHRDLAARNVLVKTQDVKITDFGLAKQLGADEKE | 293 |
| Qy | 883 | YHADGKVPITKMWALLESILRRRTHQSDVMYSGVTVVWELMTFGAKPYDGIIPAREIDPLE | 942 |
| Db | 294 | YHAEGGKVPITKMWALLESILHRIYTHQSDVMYSGVTVVWELMTFGSKPYDGIIPASEISSVLE | 353 |
| Qy | 943 | KGERLPPPICTIDVTVMYMKWMIIDSECPREFRELVSFESRMARPORFVLIQ-NEDLG | 1001 |
| Db | 354 | KGERLPPPICTIDVTVMYMKWMSDADRPKREFLIAEFSKWARPPRYLIQVQGDERMH | 413 |
| Qy | 1002 | PASPLDSTFYRSLLEDDMDGLVDABEYLVPQGGFFCPDPAPGAGGMVHRHRSSTSRG | 1061 |
| Db | 414 | LPSPDTSKFYRTLWEEDMEDIVDAEYLVPHQGGF-----NSPST--- | 454 |
| Qy | 1062 | GGDLTLGLEPSEEEAPRSP-----APSEGAGSDVFDGDLGMAAKGLOSLPHDPSPLQ | 1116 |
| Db | 455 | -----SRTPLLSLSLATSNNSATNCIDRNGG-----H----- | 481 |
| Qy | 1117 | RYSEDPVTPLPSETDGVAPLTCSPQPEYVYNQPDVRRPPSPREGPLPAARPAAGAT-LER | 1175 |
| Db | 482 | -----PVREDGFL-----PAPEYVYNQ--LMPKPKSTAMVQNIYNYISLTAISK | 523 |
| Qy | 1176 | AKTLSGKNGVVKDVFAPGGAIVENPEYL | 1203 |
| Db | 524 | LPMDSRYQN-----SHSTAVDNPISYL | 544 |

RESULT 15

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)
C.Species: avian erythroblastosis virus
C.Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 04-Feb-2000
C.Accession: B44776
R;Bruskin, A.; Jackson, J.; Bishop, J.M.; McCarley, D.J.; Schatzman, R.C.
Oncogene 5, 15-24, 1990
A.Title: Six amino acids from the retroviral gene gag greatly enhance the transforming
A.Reference number: A44776; MUID:90206603; PMID:1989616
A.Accession: B44776
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-540 <BRU>

A;Cross-references: GB:X52211
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif

Query Match 23.7%; Score 1623; DB 2; Length 540;
Best Local Similarity 54.9%; Pred. No. 6.8e-60;
Matches 340; Conservative 69; Mismatches 120; Indels 90; Gaps 14;

Qy 593 CAHYKDPPPCVARCPGSGVKPDLISYMPDKPDEEGACQPCPNCTHSCVDLDDKGCPAQ 652
Db 3 CAHFIDGPCHVKACPAAGVLGENDTL-VMKYADANAVCOLCHPNCRTGCKGPGLEGCP--- 58
Qy 653 RASPLTSIVSAVV-GILLVVLGVWFGILIKRQQOKIRKYTWRRLLQETELVEPLTPSGA 711
Db 59 NGSKTFSIAAGVVGGLCLIVVGLGIGLYLR--HIVKRTRRLRLQLRELVEPLTPSGE 117
Qy 712 MPNQAQMRLKETELRKVKVLGSAGFTGYKGIWIPDGENVKIPVAIKVLRENTSPKAN 771
Db 118 APNOAHLRLKETEFKKVKVLGFAGFTVYKGLWIPEGEKVTPVAIKELEATSPKAN 177
Qy 772 EILDENYVMAGVCSPVYSRLGICLTSTVOLITOLMPYGCLLDHVRENRGRLQSDDLNM 831
Db 178 EILDENYVMASVDNDPHVCRLGICLTSTVQLITQLMPYGCLLDYIREHKDNIGSQYLLNM 237
Qy 832 CMQIAGMSYLEDLRVLRHDLAARNVLKVSNNHVIKTFGLARLLDIDETEYHADGGVP 891
Db 238 CVQIAGMNYLEERHMVHRDLAARNVLVKTPOHVKITDFGLAKQGADEKEYHAEGGKVP 297
Qy 892 IKMALESILRRFTHQSDVWSYGVTVWELMTFGAPKDYGIPAREIPDLLEKERLPQPP 951
Db 298 IKMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEKGERLPQPP 357
Qy 952 ICTIDVYMLTWKCMIDSECRPRFRELVSFEFSMARDPORFVVIQ-NEDLGASPDLDTFF 1010
Db 358 ICTIDVYMLTWKMSGDSGRFKRELTAFFSKMARDPPRYLVIOGDERMHLPSPTDSKF 417
Qy 1011 YRSLLDDMDGLVDAEEYLV PQQGFPCPDPAAGGMVHHRRHSSTSRSGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDADEVLVPHQGFF-----NSPST----- 449
Qy 1071 PSEEAPRSL-----APSEGAGSDVFQDGLNGAAKGLQLTHDPSPQRYSBDTPVP 1125
Db 450 -----SRTPLLSSLSATSNNSATNCIDRNGG-----H----- 476
Qy 1126 LPSETDGYVAPLTCSPQEPYVNQDVVRPOPSPREGPLPAARPACAT-LERAKTISPKN 1184
Db 477 -PVREDGFL-----PAEYVNQ--LMPKKSTAMVQNQIYNIISLTAISKLPMDSRYN 527
Qy 1185 GVVKOVFAFGGAVENPEYL 1203
Db 528 -----SHSTAVDNPEYL 539

Search completed: July 22, 2003, 09:09:15
Job time : 31.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds
(without alignment)
5088.033 Million cell updates/sec

Title: SEQ4-250-264-14

Perfect score: 6839

Sequence: 1 MELAAACRWGLLLALLPPCA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 40 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| 1 | 6702 | 98.0 | 1255 | 1 ERB2_HUMAN | P06426 homo sapien |
| 2 | 5901 | 86.3 | 1257 | 1 ERB2_RAT | P06494 rattus norv |
| 3 | 5880.5 | 86.0 | 1254 | 1 ERB2_MESAU | Q60553 mesocricetu |
| 4 | 3114 | 45.5 | 1210 | 1 EGFR_HUMAN | P00533 homo sapien |
| 5 | 3093 | 45.2 | 1210 | 1 EGFR_MOUSE | Q01279 mus musculu |
| 6 | 2971.5 | 43.4 | 1308 | 1 ERB4_HUMAN | Q15303 homo sapien |
| 7 | 2952 | 43.2 | 1308 | 1 ERB4_RAT | Q62956 rattus norv |
| 8 | 2690.5 | 39.3 | 1167 | 1 XMRK_XIPWA | P13388 xiphophorus |
| 9 | 2400.5 | 35.1 | 1342 | 1 ERB3_HUMAN | P21860 homo sapien |
| 10 | 2329.5 | 34.1 | 1339 | 1 ERB3_RAT | P06279 rattus norv |
| 11 | 1922 | 28.1 | 1426 | 1 EGFR_DROME | P04412 drosophila |
| 12 | 1749.5 | 25.6 | 634 | 1 ERBB_ALV | P00534 avian leuko |
| 13 | 1703 | 24.9 | 604 | 1 ERBB_AVIER | P00535 avian eryth |
| 14 | 1630 | 23.8 | 540 | 1 ERBB_AVIEU | P11273 avian eryth |
| 15 | 1572 | 23.0 | 703 | 1 EGFR_CHICK | P13387 gallus gall |
| 16 | 1265 | 18.5 | 1323 | 1 L723_CAERL | P24348 caenorhabdi |
| 17 | 1142.5 | 16.7 | 245 | 1 ERB2_MOUSE | P70424 mus musculu |
| 18 | 733 | 10.7 | 1363 | 1 ILPR_BRALA | O02466 branchiosto |
| 19 | 721.5 | 10.5 | 1372 | 1 INSR_MOUSE | P15208 mus musculu |
| 20 | 719 | 10.5 | 1383 | 1 INSR_RAT | P15127 rattus norv |
| 21 | 716 | 10.5 | 1382 | 1 INSR_HUMAN | P06213 homo sapien |
| 22 | 699 | 10.2 | 1300 | 1 IRR_MOUSE | Q9W14 mus musculu |
| 23 | 693 | 10.1 | 1297 | 1 IRR_HUMAN | P14616 mus musculu |
| 24 | 691 | 10.1 | 1607 | 1 MIPR_LYMT | Q25410 lymphocita |
| 25 | 687 | 10.0 | 1477 | 1 HTK7_HYDAT | P14617 hydra atten |
| 26 | 686.5 | 10.0 | 1300 | 1 IRR_CAVPO | Q25197 cavia porce |
| 27 | 649 | 9.5 | 1367 | 1 IG1R_HUMAN | P08069 homo sapien |
| 28 | 636 | 9.3 | 1373 | 1 IG1R_MOUSE | Q60751 mus musculu |
| 29 | 632.5 | 9.2 | 1370 | 1 IG1R_RAT | P24062 rattus norv |
| 30 | 620 | 9.1 | 1390 | 1 INSR_AEDAE | Q93105 aedes aegypt |
| 31 | 606 | 8.9 | 2146 | 1 INSR_DROME | P09208 drosophila |
| 32 | 594.5 | 8.7 | 987 | 1 EPB4_HUMAN | P54760 homo sapien |
| 33 | 589.5 | 8.6 | 977 | 1 EPB2_MOUSE | Q03145 mus musculu |

ALIGNMENTS

RESULT 1

| ID | ERB2_HUMAN | STANDARD; | PRT; | 1255 AA. |
|----|--|-----------|------|----------|
| AC | P04626; | | | |
| DT | 13-AUG-1987 (Rel. 05, Created) | | | |
| DT | 13-AUG-1987 (Rel. 05, Last sequence update) | | | |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update) | | | |
| DE | Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112) | | | |
| DE | (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell | | | |
| DE | surface receptor HER2) (MLN 19). | | | |
| GN | ERBB2 OR HER2 OR NGL OR NEU. | | | |
| OS | Homo sapiens (Human). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | |
| OX | NCBI_TaxID=9606; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=86118663; PubMed=3003577; | | | |
| RA | Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N., | | | |
| RA | Saito T., Toyoshima K.; | | | |
| RT | "Similarity of protein encoded by the human c-erbB-2 gene to | | | |
| RT | epidermal growth factor receptor."; | | | |
| RL | Nature 319:230-234(1986). | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=86070181; PubMed=2999974; | | | |
| RA | Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A., | | | |
| RA | McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J., | | | |
| RA | Frankke U., Levinson A., Ullrich A.; | | | |
| RT | "Tyrosine kinase receptor with extensive homology to EGF receptor | | | |
| RT | shares chromosomal location with neu oncogene."; | | | |
| RL | Science 230:1132-1139(1985). | | | |
| RN | [3] | | | |
| RP | SEQUENCE OF 737-1031 FROM N.A. | | | |
| RX | MEDLINE=86016729; PubMed=2995967; | | | |
| RA | Semba K., Kanata N., Toyoshima K., Yamamoto T.; | | | |
| RT | "A v-erbB-related protooncogene, c-erbB-2, is distinct from the | | | |
| RT | c-erbB-1/epidermal growth factor-receptor gene and is amplified in a | | | |
| RT | human salivary gland adenocarcinoma."; | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985). | | | |
| RN | [4] | | | |
| RP | VARIANTS VAL-654 AND VAL-655. | | | |
| RX | MEDLINE=93194196; PubMed=8095488; | | | |
| RA | Eheanl A., Low J., Wallace R.B., Wu A.M.; | | | |
| RT | "Characterization of a new allele of the human ERBB2 gene by allele- | | | |
| RT | specific competition hybridization."; | | | |
| RL | Genomics 15:426-429(1993). | | | |
| CC | - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, | | | |
| CC | ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A | | | |
| CC | POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF- | | | |
| CC | ALPHA AND AMPHIREGULIN. | | | |
| CC | - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein | | | |
| CC | tyrosine phosphate. | | | |
| CC | - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS | | | |
| CC | (POTENTIAL). | | | |
| CC | - SUBCELLULAR LOCATION: Type I membrane protein. | | | |

P07949 homo sapien
P29317 homo sapien
Q07494 gallus gall
P54761 mus musculu
P09759 rattus norv
Q00944 gallus gall
Q91736 xenopus lae
Q91571 xenopus lae
Q91738 xenopus lae
P54762 homo sapien
P53356 hydra atten
P34152 mus musculu

| FT | DISULFID | 563 | 576 | BY SIMILARITY. |
|----------------------------|----------|-------------------------|--|--|
| FT | DISULFID | 567 | 584 | BY SIMILARITY. |
| FT | DISULFID | 587 | 596 | BY SIMILARITY. |
| FT | DISULFID | 600 | 623 | BY SIMILARITY. |
| FT | DISULFID | 626 | 634 | BY SIMILARITY. |
| FT | DISULFID | 630 | 642 | BY SIMILARITY. |
| FT | MOD_RES | 1139 | 1139 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). |
| FT | MOD_RES | 1248 | 1248 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). |
| FT | CARBOHYD | 68 | 68 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 124 | 124 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 187 | 187 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 259 | 259 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 530 | 530 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 571 | 571 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 629 | 629 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | VARIANT | 654 | 654 | I -> V. |
| FT | VARIANT | 655 | 655 | /FTID=VAR_004077. |
| FT | VARIANT | 655 | 655 | I -> V. |
| FT | CONFLICT | 1170 | 1170 | /FTID=VAR_004078. |
| FT | CONFLICT | 1255 | 1255 | P -> A (IN REF. 2). |
| FT | SEQUENCE | 1255 | 1255 | AA; 137909 MW; 39E9FDA04DCF62 CRC64; |
| Query Match | | | | |
| Best Local Similarity | | 98.0%; | Score 6702; DB 1; Length 1255; | |
| Matches 1240; Conservative | | 98.3%; | Pred. No. 0; | |
| | | | Mismatches 12; Indels 6; Gaps 1; | |
| Qy | 1 | MELAAALCRWGLLALLAPP | PGNAASTQVCTGTDMLKRLPASPTHLDMLRHLYQGCVVQGNL | 60 |
| Db | 1 | MELAAALCRWGLLALLAPP | PGNAASTQVCTGTDMLKRLPASPTHLDMLRHLYQGCVVQGNL | 60 |
| Qy | 61 | ELTYLPNLSLFLQDIQEOGVYLI | AHNOVRQVPLQRLIRVGTOLFEDNYALAVLDNG | 120 |
| Db | 61 | ELTYLPNLSLFLQDIQEOGVYLI | AHNOVRQVPLQRLIRVGTOLFEDNYALAVLDNG | 120 |
| Qy | 121 | DPLNNTTPVTGASPGGLRELQ | LSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQLA | 180 |
| Db | 121 | DPLNNTTPVTGASPGGLRELQ | LSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQLA | 180 |
| Qy | 181 | LTLLDITNRSACHPCSPMKCS | GWGESSEDCOSLTRTVCAGGCARCKGPLPTDCCHEQC | 240 |
| Db | 181 | LTLLDITNRSACHPCSPMKCS | GWGESSEDCOSLTRTVCAGGCARCKGPLPTDCCHEQC | 240 |
| Qy | 241 | AAGCTGPKHSDCLACLFHNS | HSGICELHCPALVFNNFTVSFWLRVPKVSASHLERYTFGAS | 300 |
| Db | 241 | AAGCTGPKHSDCLACLFHNS | HSGICELHCPALVYNTDTFESMINPE-----GRYTFGAS | 294 |
| Qy | 301 | CVTACPNYLYSTDVGSCTLV | CPPLHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVR | 360 |
| Db | 295 | CVTACPNYLYSTDVGSCTLV | CPPLHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVR | 354 |
| Qy | 361 | AVTSANIQEPAGCKKIFGSLA | FLPESFDGDPASNTAPLOPEQLQVETLEEITGYLIYISA | 420 |
| Db | 355 | AVTSANIQEPAGCKKIFGSLA | FLPESFDGDPASNTAPLOPEQLQVETLEEITGYLIYISA | 414 |
| Qy | 421 | WPDSLPLSVFQNLQVIRGRIL | HNHNGAYSITLQGLGISWLGSLRSLRELGSGLALIHNNTHL | 480 |
| Db | 415 | WPDSLPLSVFQNLQVIRGRIL | HNHNGAYSITLQGLGISWLGSLRSLRELGSGLALIHNNTHL | 474 |
| Qy | 481 | CFVHTVPWDLFENPHQALLH | TANRDEBCVGEGLACHOLCARGHCWGPGPTQVCNCSOF | 540 |
| Db | 475 | CFVHTVPWDLFENPHQALLH | TANRDEBCVGEGLACHOLCARGHCWGPGPTQVCNCSOF | 534 |
| Qy | 541 | LRQGEVCBEICRVLQGLPRE | VYNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDP | 600 |
| Db | 535 | LRQGEVCBEICRVLQGLPRE | VYNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDP | 594 |
| Qy | 601 | FCVARCPSGVKPDLSYMP | 1WKPPDEBGAQCPCPNCTHSCVDLDDKGCFAERASPLTSI | 660 |
| Db | 595 | FCVARCPSGVKPDLSYMP | 1WKPPDEBGAQCPCPNCTHSCVDLDDKGCFAERASPLTSI | 654 |
| Qy | 661 | VSAVVGILLVVLGVVFGIL | IKRROOKIRKYTWBRLLOETELVELPTSGAMPNQAOMRI | 720 |


```

FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD.RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD.RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SO SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.3%; Score 5901; DB 1; Length 1257;
Best Local Similarity 86.5%; Pred. No. 8e-310;
Matches 1092; Conservative 52; Mismatches 111; Indels 8; Gaps 3;

Qy 1 MELAALCRWGLLLALLPPGAASVCTCTDMKRLRPASPETHLDMRLHLYQGCQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGIAGTQCTCTDMKRLRPASPETHLDMRLHLYQGCQVQGNL 60

Qy 61 ELTYLPTNASLFLQDIOEVGYVLIHNOVROVPLQRLRIVRGTOLEFEDNYALAVLNG 120
Db 61 ELTYVPANASLFLQDIOEVGYVLIHNOVQVRLQRLRIVRGTOLEFEDNYALAVLNG 120

Qy 121 DPLNNTTPTVT-GASPGGLRELQLASLTBLKGGVLIQORNPOLCYQDTILWKDIDFHKNQL 179
Db 121 DPQDNVAASTGRTPEGLRELQLASLTBLKGGVLIQORNPOLCYQDMVLWKDVFKNQL 180

Qy 180 ALTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLETRVCAGGACRCKGPLPTDCHEQ 239
Db 181 APVDIDNRSRACPPACPKDNHCWGESPEDCQILTGTICTSGCARCKGRLPTDCHEQ 240

Qy 240 CAAGCTGPKHSDCLACLFHNSGTCCLCPALVFNFTVFWLRVPKVSASHLRYTGA 299
Db 241 CAAGCTGPKHSDCLACLFHNSGTCCLCPALVNTDTFMSMPE-----GRYTFGA 294

Qy 300 SCVTACPNYLSTDVSGCTLVCPHNOEVTAEDGTQRCCKSKPCARVCYGLGMEHLREV 359
Db 295 SCVTTCPNYLSTEVSGCTLVCPNNOEVTAEDGTQRCCKSKPCARVCYGLGMEHLRGA 354

Qy 360 RAVTSANIQEAGCKKIFGSLAFIPESDGDPAWNTAPLQEQLOVFTLEETITGLYVIS 419
Db 355 RAITSNDNVQEPDGCKKIFGSLAFIPESDGDPSGIAPLRPEQLQVFTLEETITGLYVIS 414

Qy 420 AWPDSLPLDSLVFQNLQVIRGRILHNGAYSLSLTQGLGISMGLRSRLRELGSGLALIHNT 479
Db 415 AWPDSLRLDSLVFQNLRIIRGRILHNGAYSLSLTQGLGISHLSRLRELGSGLALIHNAH 474

Qy 480 LCFVHTVPWDQLFRNPHQALLHTANRPEDE-CVSEGLACHOLCARGHCWGPGPTQCVNCS 538
Db 475 LCFVHTVPWDQLFRNPHQALLHSGNRPEDELCVSSGLVCNLSLCAHCHCWGPGPTQCVNCS 534

Qy 539 QFLRGQECVSCRVLQGLPRYVNAHCLCPCHPECPQNGSVTCFPGPADQCVCAHYKD 598
Db 535 HFLRGQECVSCRVMKGLPREYVSDKRCCLPCHPECPQNSSETCFGSEADQCAACHYKD 594

Qy 599 PPFCVAPCPGVKPDLSWMLTWKFPDEGACQPCPINCTHSCVDLDKDCGPAEORASPLT 658
Db 595 SSSCVAPCPGVKPDLSWMLTWKFPDEEGICQPCPINCTHSCVDLDKDCGPAEORASVPT 654

Qy 659 SIVSAVVGILLVWVVGILIKRRQOKIRKYMRLLOETELVEPLTPSGAMPNQAM 718
Db 655 FIATVVGILLVWVVGILIKRRQOKIRKYMRLLOETELVEPLTPSGAMPNQAM 714

Qy 719 RILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAY 778

```

```

Db 715 RILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAY 774
Qy 779 VMAGVSPYVSRLLIGICLTSTVOLVQLMPYGCCLLDHVRENRRGRSGODLLNMCQIANG 838
Db 775 VMAGVSPYVSRLLIGICLTSTVOLVQLMPYGCCLLDHVREHRRGRSGODLLNMCQIANG 834
Qy 839 MSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGCKVPKIMMALE 898
Db 835 MSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGCKVPKIMMALE 894
Qy 899 SILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVY 958
Db 895 SILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVY 954
Qy 959 MIMVKCMIDSECRPRRELSEFSRWARDQRFVVIQNEIDLGPASPLDSTFYRSLLDD 1018
Db 955 MIMVKCMIDSECRPRRELSEFSRWARDQRFVVIQNEIDLGPSPMDSTFYRSLLDD 1014
Qy 1019 DMGDLVDAEYLVPOQGFPCDPAPGAGMWHHRHRSSTRSGGDLTLGLPEPSEEAAPR 1078
Db 1015 DMGDLVDAEYLVPOQGFPCDPPTGCTGTAHRRHRSSTRSGGDLTLGLPEPSEGP 1074
Qy 1079 SPLAPSEGAGSDVDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLT 1138
Db 1075 SPLAPSEGAGSDVDGDLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLA 1134
Qy 1139 CSPOPEYVQNPVDRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVVDYPAFGGAVE 1198
Db 1135 CSPOPEYVQNPVDRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVVDYPAFGGAVE 1194
Qy 1199 NPEYLTPOGGAAPPHPPAFSPAFDNLVYDQDPPRPAPESTFGTPTAENPEYVLGD 1258
Db 1195 NPEYLVREGTASPHPSPAFPAFNLVYDQNSSEGOGPPSPNFEPTAENPEYVLGD 1254
Qy 1259 VPV 1261
Db 1255 VPV 1257

RESULT 3
ERB2 MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERBB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RA MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

```


OY 1021 GDLVDAEYLVPPQGFCCPDAPGAGMVGHHRRSSSTRSGGDLTLGLEPSEEEPRSP 1080
|||||
DB 1015 GDLVDAEYLVPPQGFCCPDAPGAGMVGHHRRSSSTRSGGDLTLGLEPSEEEPRSP 1074
|||||
OY 1081 LAPSEAGSDVPDGLGMAKGLQSLPHTHDSPLQRYSEDTVPPLSETDGYVAPLTCS 1140
|||||
DB 1075 LAPSEAGSDVPDGLGMAKGLQSLPHTHDSPLQRYSEDTVPPLSETDGYVAPLTCS 1134
|||||
OY 1141 POPEVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPKNGVVDVAFAGAVENP 1200
|||||
DB 1135 POPEVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPKNGVVDVAFAGAVENP 1194
|||||
OY 1201 EYLTPQGAAPQPPPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYILGLDVP 1260
|||||
DB 1195 EYLTPQGAAPQPPPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYILGLDVP 1253
|||||
OY 1261 V 1261
DB 1254 V 1254
RESULT 4
EGFR_HUMAN
ID EGFR_HUMAN STANDARD; PRT: 1210 AA.
AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;
AC O00688; Q9BZS2; Q9H2C9; Q9GZX1; Q9H3C9;
DT 01-JUN-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
protein-tyrosine kinase ErbB-1).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL Nature 309:418-425(1984).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=95382957; PubMed=7654368;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta.";
RL Mol. Reprod. Dev. 41:149-156(1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Maible N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RT factor receptor gene encodes a truncated form of the receptor.";
RL Nucleic Acids Res. 24:4050-4056(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97256547; PubMed=9103388;
RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT "Expression of a truncated epidermal growth factor receptor-like
RT protein (TEGR) in ovarian cancer.";
RL Gynecol. Oncol. 65:36-41(1997).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC TISSUE=Placenta;
RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGFR transcripts encoding truncated receptor
RT isoforms.";
RL Genomics 71:1-20(2001).
RN [6]
RP SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Human and mouse alternative EGFR transcripts encoding only the
RT extracellular domain of the receptor.";
RL Submitted (FEB-1999) to the ENBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells.";
RL Science 224:843-848(1984).
RN [8]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245935; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells.";
RL Nature 309:806-810(1984).
RN [9]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells.";
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN [10]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription.";
RL Oncogene Res. 1:375-396(1987).
RN [11]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis.";
RL J. Biol. Chem. 266:1746-1753(1991).
RN [12]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [13]
RP SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [14]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor


```
Oy 965 WMIDECRPRRELVSFMRWDPQRFVVTQ-NEDLGPAASPLDSTFYRSLLEDDMGDL 1023
Db 951 WMIDADSPKRELIIKFSKMDPQRYLVITQGDHMLPSPDTDSNFYRALMDEEDMDV 1010
Oy 1024 VDAEYLVPQGFPCPDPAAGAGVHRRSSSTRSGGDLTGLPSESEAEPRSLAP 1083
Db 1011 VDAEYLVPQGF-----SSPSTGRTPLLS 1036
Oy 1084 SEGAGSDVFDGLGMAKGLQSLTHDPSPQLQRYSEDTPVLPSET--DGVPAPLTCS 1141
Db 1037 SLSATSN--NSTVACIDRNGLSQCPKEDSFQRYSSDPTGALTEDSIDDTFL-----P 1088
Oy 1142 QPEYVQSDVRQPPSPREGPLPAPRAGATLERAKTLSPKNGVVKDVFAGGAVENPE 1201
Db 1089 VPEYVINO-SVKRPAGSVQNVYHNPQPLNP-----APSRDPHYQD--PHSTAVGNPE 1137
Oy 1202 YL-TPOGAAQPPPPPAFAFNLYYWDQ-----DP-----PERGAPPSTFK 1244
Db 1138 YLNTVQ-----PTCVNSTFDSPAHMAQKSHQISLDNPVQDDFPFKPAKPNIGFK 1188
Oy 1245 GTPTAENPEYL 1255
Db 1189 GS-TAENAEYL 1198

RESULT 5
EGFR_MOUSE
ID EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC 001279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
(1)
SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RL (bek/KGFR) gene."
RL Oncogene 7:1957-1962(1992).
(2)
SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RL in mouse blastocysts during delayed implantation."
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
(3)
SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs M.L.;
RT Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
(4)
SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RL receptor tyrosine kinase."
RL Genes Dev. 8:399-413(1994).
(5)
SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
```

```
RT "Comparison of EGF receptor sequences as a guide to study the ligand
binding site."
RL Oncogene 6:673-676(1991).
(6)
SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eisinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (A PROTEIN TYROSINE KINASE).
CC !- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC !- SUBCELLULAR LOCATION: Type I membrane protein.
CC !- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC !- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
-----
EMBL; X78987; CAA55587.1; -
EMBL; U03425; AAA17899.1; -
EMBL; X59698; CAA42219.1; -
EMBL; L06864; AAA53029.1; -
EMBL; Z12608; CAA78249.1; -
HSSP; P11362; IFGK.
MGP; MGI:95294; Egfr.
InterPro; IPR000494; EGFR_L_domain.
InterPro; IPR000719; Euk_pkinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR01245; Tyr_pkinase.
Pfam; PF00069; Pkinase; 1.
Pfam; PF00757; Furin-like; 1.
Pfam; PF01030; Recep_L_domain; 2.
ProDom; PD000001; Euk_pkinase; 1.
SMART; SM00261; FU; 3.
SMART; SM00219; Tyrc; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
SIGNAL 1 24
FT CHAIN 25 1210 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 25 647 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 648 670 POTENTIAL.
FT DOMAIN 671 1210 CYTOPLASMIC (POTENTIAL).
FT REPEAT 75 300 APPROXIMATE.
FT REPEAT 390 600 APPROXIMATE.
FT DOMAIN 1028 1071 SER-RICH.
FT DOMAIN 714 981 PROTEIN KINASE.
FT NP_BIND 720 728 ATP (BY SIMILARITY).
FT BINDING 747 747 ATP (BY SIMILARITY).
FT ACT_SITE 839 839 BY SIMILARITY.
FT DISULFID 194 199 BY SIMILARITY.
FT DISULFID 207 207 BY SIMILARITY.
FT DISULFID 215 223 BY SIMILARITY.
FT DISULFID 219 231 BY SIMILARITY.
FT DISULFID 232 240 BY SIMILARITY.
FT DISULFID 236 248 BY SIMILARITY.
FT DISULFID 251 260 BY SIMILARITY.
FT DISULFID 264 291 BY SIMILARITY.
FT DISULFID 295 307 BY SIMILARITY.
FT DISULFID 311 326 BY SIMILARITY.
FT DISULFID 329 333 BY SIMILARITY.
```

```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
FT SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

```

Query Match 45.2%; Score 3093; DB 1; Length 1210;

Best Local Similarity 49.1%; Pred. No. 6e-159;

Matches 628; Conservative 171; Mismatches 363; Indels 116; Gaps 24;

```

Qy 11 LLLALLPPGA--STQVCTGTGMKRLPASPETHLDMLRHLYQCGVQVQGNLEITYLPTN 68
Dy 14 LTLAALCAAGGALEKKVCCQSTNSRLTQLGTGFEDHFLSLQRMYNCEVVLGNLEITYVQRN 73
Qy 69 ASLSFLQDIQSVQVGLIAHNOVQLQRLIRVGTQFEDNYVALALDNGDPLNNTTP 128
Dy 74 YLSFLKTIQEVAGVLIANTVERIPLENQIIRGNALYENTYALATLSN----- 124
Qy 129 VTGASPGGLRELQRLSLTEILKGGVLIQRNPLQCYQDTILWKDI----FHKNNOLALTLI 184
Dy 125 -YGNTRGLRELPMRNLEILLIGAVRFSNNPILCNMDTIQWRDIQNVFMSNMSMDL--- 180
Qy 185 DTRNSRACHPCSPMKCSRGWSESDCQSLTRTVACGGCA-RCKGPLPTDCHEHQCAAG 243
Dy 181 -QSHPSKCPKCDPSCPNCSGCGGGEENCQKLTKIICAQCCSHRCRGRSPSDCHNQCAAG 239
Qy 244 CTGPKHSDCLACHFNHSGICELHCPALVFNFTVSEFWLRVPKVSASHLERYTFGACVT 303
Dy 240 CTGPRESDCLVCKQFQDEATCKDTCPLMLNPTY-----QMDVNPGEKYSFGATCVK 293
Qy 304 ACPVNYLSTDVGSCTLVCPHLNQBVTABDGTQRCCKSPKCARVCYGLGMEHLREVRVAT 363
Dy 294 KCPNVYVTDHGSVCVRAGDPDYEV-BEDGIRKCKCDGPKCKVCNGIGIGEFKDTLSIN 352
Qy 364 SANIQEAGCKKIFGSLAFIPESFDGPASNTAPLOPELOVFTLBEITYLISAWPD 423
Dy 353 ATNIKHFKYCTAISGDLHILPVAEFGDSFTRTPPLPRELEILKTVKEITGFLLIQAWPD 412
Qy 424 SLPDLVSFQNLQVIRGRILHNGAYSLTQIGISWLGRLSRLRELGLALIHNTLHLCFV 483
Dy 413 NWTDLHAFENLEIIRGTQKHQQLQSLVAVGLNITSLGLRSLKEISDGDVIISGNRNLCTYA 472
Qy 484 HTVPWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCKWGPQTQCVNCSQFLRG 543
Dy 473 NTINWKKLFGTPNQKTKIMNRAEKCKAVNHNVCNPLCSGCGKGPFRCDVCSQNVSRG 532
Qy 544 QECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCACAHYKDPFPVCV 603

```

```

Db 533 RECVEKCNILEGEPRFVENSECICQHPCELPQAMNITCTGRGPNOCIQCAHYIDGPHCV 592
Qy 604 ARCSGVKPDLSYMPIWKFPDEEGACQCPINCTHSCVDLDDKGCFAEORASPLTSIVSA 663
Db 593 KTCFAGINGENNTL--VMKYADANNVCHLCHANCTYGCAGPGLOQCEVWSPGPKIPSIATG 651
Qy 664 VVGILLVVVLGVVFGI-LIKERQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAORILK 722
Db 652 IVGGLLFIVV-VALGIGLPMRRRHIVRKTLLRLLQERELVEPLTPSGEAPNQAHLILK 710
Qy 723 ETELKVKVLGSGAFGTYYKGIWIPDGENKIPVAIKVLRENTSPKANKEILDEAYVMAG 782
Db 711 ETEFKIKIVLGSAGFTYVKGWIPEGEKVIPVAIKELREATSPKANKEILDEAYVMAS 770
Qy 783 VGSPPVSLGICLTSTVOLTPMLPYGCLLDHVRNRLGSOALLNWCMIQAKGMSYL 842
Db 771 VDNPHVCRLLGLCTSTVOLITQMLPYGCLLDYVREHKDNGISQYLLNWCVIQAKGMNLY 830
Qy 843 EDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMALESILR 902
Db 831 EDRRLVHRDLAARNVVKTPQHVKITDFGLAKLLGAEEKYHAEKGKVPKIKMALESILH 890
Qy 903 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIW 962
Db 891 RIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASDISSILEKGERLPQPPICTIDVYMIW 950
Qy 963 KCMWIDSECRFRFRELSEFSEPMARDPQRFVQI-NEDLGPASPLDSTFYRSLLEDDMG 1021
Db 951 KCMWIDADSRPKFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEME 1010
Qy 1022 DLVDAEYLVPOQGFCCPDPAAGAGVMVHRHRSSTRSGGDLTLGLLEPSEEAAPRSP 1081
Db 1011 DTVDAEYLVPOQGF-----NSPST-----SRTP 1036
Qy 1082 APSEGAGSDVFDGLGMGAAGKLOSLPHDPSPLQRYSEDTVPPLSET--DGYVAPLTC 1139
Db 1037 LSSLSATSN---NSTVACINRNGSCRVKEDAFLOQYSSDPTGAVTEDNIDDAFL---- 1087
Qy 1140 SPOEYVYNQPVVRQPPSPREGPLPAAPACATLERAKTLSPKNGVGVKDVFAFGAVEN 1199
Db 1088 -PVPEYVYNQ-SVPKRPAGSVQNVYHNQPLHP-----APGRDLHYQN--PHSNVAVGN 1135
Qy 1200 PEYL-TPQGGAAPHPHPPAPAFDNLVYWDQ-----DP-----PERGAPPST 1242
Db 1136 PEYLNQA-----PTCLSSGFNSPALMIQSHQMSLDNDPDYQDFFPKETKPNGI 1186
Qy 1243 PKGTPTAENPEYLGLDVP 1260
Db 1187 FKQ-PTAENAEYLRVAP 1203

```

RESULT 6

```

ID ERB4 HUMAN STANDARD; PRT; 1308 AA.
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
GN (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
OS ERBB4 OR HER4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the

```

epidermal growth factor receptor family.";
Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
[2]
SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
TISSUE=Fetal brain;
MEDLINE=97476287; PubMed=9334263;
Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
Klagsbrun M.;
A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
tissue distribution and differential processing in response to
phorbol ester.";
J. Biol. Chem. 272:26761-26768(1997).
CC -1- NRG-3 SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
2. NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
NON-CLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
PLUTARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
LUNG, SALIVARY GLAND, AND PANCREAS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; L07868; AAB59446.1; -;
DR HSP; P11362; 1FGK.
DR Genew; HGNC:3432; ERBB4.
DR MIM; 600543; -;
DR InterPro; IPR0000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TVR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
KW SIGNAL
FT CHAIN 1 25
FT CHAIN 26 1308
FT DOMAIN 26 651
FT TRANSMEM 652 675
FT DOMAIN 676 1308
FT DOMAIN 186 334
FT DOMAIN 496 633
FT DOMAIN 718 985
FT DOMAIN 985
FT PROTEIN KINASE.
FT RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT CYS-RICH.
FT CYS-RICH.
FT PROTEIN KINASE.

```
Db 349 DSSNIDKFINCTKINGNLIFLVTGIGHDPYNAIEADPEKLNVRTVREITGFLNIQSWP 408
Qy 423 DSLPDLVSFQNLQVIRGIRLHNGAYSLSLTQGLGISWLGRLSRLRELGSLALHTHNLCHF 482
Db 409 PNMTDFSVSFLNLTIGGRVLSGLSLILKQGITSLQFQSLKEISAGNIYITDNSLCY 468
Qy 483 VHTVPWDLFRNPQHALLHTANRPEDECVGEGLACHQICARGHCWGPPTQCVCNSQFLR 542
Db 469 YHTINMTLFTSTINRIIVIRNKAENCTAEGMVCNHLCSGDCGPGPDQCLSCRRFSR 528
Qy 543 GOECVEECRVQLGLPREVYNARHCLUPCHPEOP-ONGSVTCFGEADOCVACAHYKDDPF 601
Db 529 GRICIESCNLDGFEREFENGISICECDPQCEKEMEDGLTCHGPGPDNCTCKSHFDQGN 588
Qy 602 CVARCPGVKPDLSWMPYIWKFPDEEGACOPCPINTHSCVDLDDKC-----PA 650
Db 589 CVEKCPDLOGANSF--IFKVADPDRECHPCNPCTGCGNGPTSHDCIYVPMWGHSTLQP 646
Qy 651 EORASPLTSIVSAV-GILLVVVLGVFGILIKRQOKIRKYTWRLLOETELVEPLTPS 709
Db 647 HAR-TPL--IAAGVIGGUFILVGLTFVAVYRRKSIK-KKRALRRFL-ETELVEPLTPS 701
Qy 710 GAMPNQAQMRILKTELKRVKVLGSCAFGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKA 769
Db 702 GTAPNQAQRLILKTELKRVKVLGSCAFGTGVYKGIWVPEGETVKIPVAIKILNETTGPKA 761
Qy 770 NKEILDEAYMAGVSPYVRLGLICLTSTVQLVTQLMPYGCLLDHVRENRRGLSGDQL 829
Db 762 NVEFMDEALIMASDHPHLVRLGLVCLSTPTQLVTQLMPHGCCLLEYVHEHKDNIGSQLL 821
Qy 830 NWKQIAGMSYLEVDRLVRLHDLAARNVLKSPNHVKITDGLARLLDIDETEHADGCK 889
Db 822 NWKQIAGMNYLERRLVRLHDLAARNVLKSPNHVKITDGLARLLLEGEKEYNADGCK 881
Qy 890 VPIKWMALLESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGPICAPAREIPDLLEGERLPQ 949
Db 882 MPKWMALCETHYKFTHQSDVWSYGVTVWELMTFGGKPYDGPITREIPDLLEGERLPQ 941
Qy 950 PPICTIDVYIMVKCWMIDSECRPRPRELVSEFSRMARDPQRFVVIQNEED-LGPASPLDS 1008
Db 942 PPICTIDVYIMVKCWMIDADSRLPKFKELAAEFSSRMARDPQRYLVIQGDDRMKLPSPNDS 1001
Qy 1009 TFYRSLLEDMDGLVDAAEYLVVQOQFCFDPAPGACGMVHRRHSSTSGGDLTLG 1068
Db 1002 KFFONLDEEDLMDMAEYLVLP-QAFNIPPP-----IYTSRARIDSNRS-----EIG 1049
Qy 1069 LEPSEEEAPRS-----PLAP-SEGAGSDVFDGDLGMG 1099
Db 1050 HSPPPAYTPMSGNQFVYRDGFAAEQGVSVYPRAPTSTIPEAPVAQGTAFIFDDSCNG 1109
Qy 1100 AAKGLQSLPTHDPSPLOQRYSEDPTVPLPS-----ETDGYVAPLTCSPQPEYVNPQDVR 1152
Db 1110 TLRKPVAPHVQEDSSTORYSADPTVFAPERSPRGELDEGYMTWRDKPKQEYLNPNVE-- 1167
Qy 1153 PQPSPREGPLPAARPAGATLERAKLTLSPGNGVVKVDFAGGAVENPEYLTPOGGAAPQ 1212
Db 1168 -----ENPVSRR-----KXGDLQ-----ALDNPYHNASNG----- 1194
Qy 1213 PHPPPA-----PSPAFDNLVYWDQDPPPERGA 1238
Db 1195 --PKADEYVNEPLYLNTFANTLKGAEYLNKNTLSMPEKAKKAFDNDPYWNHSLPRST 1252
Qy 1239 --PPSTFKGTPT-----AENPEYL 1255
Db 1253 LQHPDYLQYESTKYFKYQNGRIRPIVAENPEYL 1285

RESULT 7
ERR4 RAT
ID ERR4 RAT
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
```

```
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RJ [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RJ [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RJ J. Neurosci. 17:1642-1659(1997).
CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3. HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN. WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALLAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF041838; AAD08899.1; -
EMBL; U52531; AAC53051.1; -
HSSP; P11362; 1FGK.
InterPro; IPR000494; EGFR_L domain.
InterPro; IPR000719; Euk_Pkinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR001245; Tyr_Pkinase.
InterPro; IPR004019; YLP motif.
Pfam; PF00757; Furin-like; 1.
Pfam; PF00069; Pkinase; 1.
Pfam; PF01030; Recep_L_domain; 2.
Pfam; PF02757; YLP; 2.
PRINTS; PR00109; TYRKINASE.
ProDom; PD000001; Euk_pkinase; 1.
```

DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TyRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 995 PROTEIN KINASE.
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PXT -> STR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match 43.2%; Score 2952; DB 1; Length 1308;
Best Local Similarity 45.1%; Pred. No. 2 5e-151;
Matches 610; Conservative 194; Mismatches 382; Indels 168; Gaps 30;

QY 1 MELA-ALCRWGLLL--ALLPQGAATQVCTGDMKRLRPLASPETHLDMLRHLYQGCQVVQ 57
DB 1 MKLATGLWVGSLVAARTVQPSASQSVCAGTENKLSLSLDLEQQVRLRKYENCEVVM 60
QY 58 GNLEITYLPTNASLFLQDIQSVGVLTAAHNOVQVPLORIRIVRGTLFEDNYALVL 117
DB 61 GNLEITSYHNKDLFLRSIREVTGYVLVALNQFRLPLENRIIRGTGKLYEDYALAI 120
QY 118 DNGDPLNNTPTVTGASPGGLRELQLRLSLEILKGVLIQPNPOLCVQDTILWKDIFHKNN 177
DB 121 LNYRKDGNF-----GLQELGLKNLLEILNGGVYVDQNKFLCYADTIHQWDIRNPW 171

QY 178 QLALTLIDNRSRACHPCSPCKGRCWGESSEDCQSLTRTVCAAGC-ARCKGPLPTDCC 236
DB 172 PSNMTLVSTIGSSGCRCHKSTG-RCWGPTENHCOTLRTVCAECCDRCYGPYVSDCC 230
QY 237 HQCAAGCTGPKHSCLACLHPNHSGICELHCP-ALVFNFTVSVLWLVKVSASHLEY 295
DB 231 HREACAGCGSGPKDTCFACMFNDSGACVTCQTPQTFVYNPTTF-----QLEHNFNAKY 283
QY 296 TFGASCVTACPNYLSTVGSCVTLVCPLHNQVTAEDGTQRCCKSKPCARVCYGLGMEH 355
DB 284 TYGAFCVKCPHNFV-VDSSSCVRACPSKMEV-BENGIMCKPCPTDIPCACDGI GTGS 341
QY 356 LREVAVTSANIQEFAGCKKIFGSLAFIPESPDGPASNTAPLQPELOLVFETLEBITGY 415
DB 342 LMSAQTVDSNIDKFINCTKINGNLIFLVTHGDPYNAIDAI DPEKLVNFRVREITGF 401
QY 416 LVISAWPSLPDLVSFQNLQVIRGRILHNGAYSLTQGLGIGLWLSRLSRELGSGALAIH 475
DB 402 LNIQTWPPNMTDFSVFNSLVITIGRVLVSGLSLLILKQOGITSLQFSLKEISAGNIYIT 461
QY 476 HNTHLCFVHTVPMDOLFNPHQALLHTANRPEDECVGEGCLACHOLCARGCHWCPGTQCV 535
DB 462 DNSNLCYHTINMTTLFSTVNQRIVIRDNRAENCTAEGMVCNHLCSNDCGCGPGDQCL 521
QY 536 NCSQFLRGQECVECKRVLQGLPREYVNAHCLPCHPECOF-QNGSVTCFGEADQCVACA 594
DB 522 SCRRFSRGKICIESCNLDYGEFREFENGSI CECDSQCEKMEDEGLTCHGPGDNCCKS 581
QY 595 HYKDPFFCVARCPGKVPDLSYMPIWKFPDEGACQPCPINCTHSCVLDLDDKGC----- 648
DB 582 HFKDGPNCVEKCPVLQGANSP--IFKYADQRECHPCHPNCTQCGNCTSHDCIYVPWT 639
QY 649 -----PASORASPLTSIYSAVV-GILLVVVLGVGVGILLKRRQOKIRKYMRLLOTEL 702
DB 640 GHTSLPQHAR-TPL--IAAGVIGGLFVIMALTFAVYVRRKSIK-KGALRRFL-ETEL 694
QY 703 VEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVL 762
DB 695 VEPLTPSGTAPNQAQRLILKETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKILN 754
QY 763 ENTSKANKEILDEAYVWAGVSPVSRLLGICLTSTVOLVTOLMPYGLCLDHLVRNRR 822
DB 755 ETTGPKANVEFMDEALINASVDHPLVRLGLVCLSPITQLVTLQMLPHGCLLYVHEHKN 814
QY 823 LGSQDLNWKQIAKMSYLEVDVRLVHRDLAARNVLKSPNKHVKTDFGLARLLDDETE 882
DB 815 IGSQDLNWKVQIAKGMYLEERRLVHRDLAARNVLKSPNKHVKTDFGLARLLDEGEKE 874
QY 883 YHADGGKVPIMWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLE 942
DB 875 YNADGGKVPIMWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLE 934
QY 943 KGERLPQPPICITIDVYIMVKWCMIDSECRPRFRELVSFSEFMSHARDPQRFVIONED-LG 1001
DB 935 KGERLPQPPICITIDVYIMVKWCMIDSECRPRFRELVSFSEFMSHARDPQRFVIONED-LG 994
QY 1002 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCDDP----- 1041
DB 995 LPSPNDSRFQNLLEDEEDMDAEYLVLP-QAENIPPIYTSRTRIDSNRSEIGHSP 1053
QY 1042 ---AFGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEAGSVDVFGDGLGM 1098
DB 1054 PAYTPMSGSQFVYQDGFATQGG---MPMPYATTSTIPEAPVA--QGATAEMFDDSCCN 1108
QY 1099 GAAGLQSLPHDPSFLORYSEDPTVPLPS-----ETDGVVAPLTCSQPEYVNPQDV 1151
DB 1109 GTLRKPVVPHVQEDSSTQRYSDPTVFAPERNPRAELDBEGYMTMHDKPKQEYLNAPVE- 1167
QY 1152 RPQPSPREGLPAARPAGATILERAKTLSPGKNGVVKDFVAFGGAVENPEYLTPOGGAAP 1211
DB 1168 -----ENPFVSR-----KNGDLQ-----ALDNPYHSASSG----- 1194
QY 1212 QPHPPA-----FSPAFDNLVYWDQDPPERG 1237

Db 1195 ---PPKAEDEVNPEPLYLNTFTNALGNAEYMKNSLLSVPEKAKAFDPNDYVHNSLPDRS 1251
 Qy 1238 A--PPSTFKGTPT-----AENPEYL 1255
 Db 1252 TLQHPDYLQESTYKFKYKQNGRIRPIVAENPEYL 1285

RESULT 8
 XMWK_XIPMA STANDARD; PRT; 1167 AA.
 ID XMWK_XIPMA AC P13388;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMWK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Wittbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,
 RA Telling A., Robertson S.M., Scharlt M.,
 RA "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RT inducing Tu locus in Xiphophorus.";
 RL Nature 341:415-421 (1989).
 RN [2]
 RP REVISION TO 515.
 RA Schardt M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X16891; CAA34770.2; -
 DR PIR; S06142; S06142.
 DR HSPG; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR022290; Ser_thr_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00220; S_TKC; 1.
 DR SMART; SM00219; TyrKC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
 FT FT

Db 404 EITGLVIMWNNPNTSLVFNLEIIRGRTTFGRGFSFVVVQVRHLQWLGRLSLKEVSA 463
Qy 470 GLALIHNTLHLCFVHTVPMQDLFNPQALHHTANRPEDECVEGELACHQICARGHCWP 529
Db 464 GNVILKNTLQRYANTINWRFLRSEDOSIEYDART-----ENQTCNNECEDGCGWP 516
Qy 530 GPTQCVNCSQPLRGOECVEECRVLQGLPREVYNARHCLPCHPEQOPQNGSVTCFGEADQ 589
Db 517 GPTCVNCSLHVDGRGRCVASCNLLQGEPREAQVDCRCVQCHQECVLQVDTSLTCYGPAN 576
Qy 590 CVACAHYKDPFCVACRSGVKPDLSPYINPKFDPDEBAGCQPCINCTHSCVDLDDKCP 649
Db 577 CSKSAHFQDGPQICPRCPHGILGDDTL-IWKYADKMGQCQPCQNCQTCGCGSGLSGR 635
Qy 650 AEORASPLTSIVSAVVGILLVVLGWFGLIKRRQOKIRKYTWRRLLQETELVEPLTPS 709
Db 636 GD-IVSHSLAVGLVSGLLITVIVALLVLLRRRIK-RKRTIRCLLQELVELVEPLTPS 693
Qy 710 GAMPNQAOMRILKETELRKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRRENTSPKA 769
Db 694 GOAPNAQFLRIKTEFKKRVLGSGAFGTVYKGLWNPEDGENIRIPVAIKVLRRENTSPKV 753
Qy 770 NKEILDEAYNAGVSPVSRLLGLICTSTVQLTQMLPYGCLLDHVRNRRGLSGDQL 829
Db 754 NOEVLDEAYNMAVDHPHVCRLGLICTSAVQLTQMLPYGCLLDYVVRQHOERICQWLL 813
Qy 830 NWCQIAGMSYLEDVRLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGK 889
Db 814 NWCQIAGMYLEERHLVLRDLAARNVLLKPNHVKITDFGLSKLTADKEVQADGK 873
Qy 890 VPIKWMALLESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQ 949
Db 874 VPIKWMALLESILQWYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPQ 933
Qy 950 PPICTIDVYMTVMKWMIDSECRPRFRELVEFSFMAARDPQFVVVQIENEDLGPASPLDST 1009
Db 934 PPICTIEVYMTILKWMIDSPSRFRFRELVEFSQMARDPKRYLVIQG---NLPSLSDRR 990
Qy 1010 FYRSLLEDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTGL 1069
Db 991 LFSRLSDD--DVDADEYLLPYKRI-----NRQGS----- 1020
Qy 1070 EPSEEAAPRSLAPSEGAGSDVFDGLGMAKGLQSLTPHDPSPQLQRYSEDPTV-PLPS 1128
Db 1021 -----EPCIPPTG-----PVRENSITLRNISDPTQNALEK 1051
Qy 1129 ETDGYVAPLTCSPQPEYVQDPVRQP-----PSPRE-----GPLP-AARPAGATLER 1175
Db 1052 DLGDH-----EYVNPQGETSSRLSDIYNPNYEDLTDGMPVSLSQEAETNFSR 1101
Qy 1176 AKTLSPGKGVVQVFAFGGAVENPEYLTPOGGAAPQHPFPATSPAFDNLVYWDQDPE 1235
Db 1102 PEYLTNQNSI---PLVSSGSMDDPY---QAG-----YQAAE-----LPQ 1136
Qy 1236 RGAPPSTFKGTPTAENPEYLG 1256
Db 1137 TGALTNGMFLPAAENLEYLG 1157

RESULT 9

ERB3_HUMAN
ID ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (tyrosine kinase-type cell surface receptor HER3).
GN ERBB3 OR HER3
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9008234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RX TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -|- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -|- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -|- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -|- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -|- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -|- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -|- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; M29366; AAA35790.1; -;
CC EMBL; M34309; AAA35979.1; -;
CC EMBL; S61953; AAB26935.1; -;
CC PIR; A36223; A36223.
CC HSSP; P11362; 1FGK.
CC Genew; HGNC:3431; ERBB3.
CC MIM; 190151; -;
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recept_L domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

| | | |
|---|-----------------------|---|
| KW | Alternative splicing. | |
| FT | SIGNAL | 1 19 |
| FT | CHAIN | 20 1342 |
| FT | DOMAIN | 20 643 |
| FT | TRANSMEM | 644 664 |
| FT | DOMAIN | 665 1342 |
| FT | DOMAIN | 709 966 |
| FT | NP_BIND | 715 723 |
| FT | BINDING | 742 742 |
| FT | ACT_SITE | 834 834 |
| FT | DISULFID | 186 194 |
| FT | DISULFID | 190 202 |
| FT | DISULFID | 210 218 |
| FT | DISULFID | 214 226 |
| FT | DISULFID | 227 235 |
| FT | DISULFID | 231 243 |
| FT | DISULFID | 246 255 |
| FT | DISULFID | 259 286 |
| FT | DISULFID | 290 301 |
| FT | DISULFID | 305 320 |
| FT | DISULFID | 323 327 |
| FT | DISULFID | 500 509 |
| FT | DISULFID | 517 517 |
| FT | DISULFID | 520 529 |
| FT | DISULFID | 533 549 |
| FT | DISULFID | 552 565 |
| FT | DISULFID | 556 573 |
| FT | DISULFID | 576 585 |
| FT | DISULFID | 589 610 |
| FT | DISULFID | 613 621 |
| FT | DISULFID | 617 629 |
| FT | CARBOHYD | 126 126 |
| FT | CARBOHYD | 250 250 |
| FT | CARBOHYD | 353 353 |
| FT | CARBOHYD | 408 408 |
| FT | CARBOHYD | 414 414 |
| FT | CARBOHYD | 437 437 |
| FT | CARBOHYD | 469 469 |
| FT | CARBOHYD | 522 522 |
| FT | CARBOHYD | 566 566 |
| FT | CARBOHYD | 616 616 |
| FT | VARSPLIC | 141 183 |
| FT | VARSPLIC | 184 1342 |
| FT | CONFLICT | 560 560 |
| FT | CONFLICT | 1064 1064 |
| SEQ | SEQUENCE | 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64; |
| Query Match 35.1%; Score 2400.5; DB 1; Length 1342; | | |
| Best Local Similarity 40.3%; Pred. No. 1.1e-121; | | |
| Matches 531; Conservative 193; Mismatches 457; Indels 137; Gaps 34; | | |
| Qy | 10 | GLLALLPPGAA--STOVCTGDMKRLPASPETHLMDRLHLYOGCOVVOGNLELTLYLPT 67 |
| Db | 11 | GLFLSGLARGEVNSQAVCGFTLNGLSVTGDAENQYQTLKYLERCEVVMGNLEIVLTGH 70 |
| Qy | 68 | NASLSFLQDIOEVGYVLIHNRQVQLRQLRIRVGTQLFEDNYALAVLDNGDPLNNTT 127 |
| Db | 71 | NADLSFLQIREVTGYVLVAMNEFTLPLNLRVVRGTVQYDGFVFM-----LNYNT 125 |
| Qy | 128 | PVTGASPGGRLRLRLSLILKGVLIQRNPQLCYODTILWKDIFHNHQLALTLDTN 187 |
| Db | 126 | -----NSSHALRQLRLTQTLISGGVYIEKNDKLCHEMDTIDWRDIVDRD---AEIVVKD 178 |
| Qy | 188 | RSRACHPCSPMKGSRGCSSESDCSLIRTVACAGC-ARCKGPLPTDCHEQCAACTG 246 |
| Db | 179 | NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNHGCFGNPNQCCHDECAGCSG 237 |
| Qy | 247 | PKHSDCLACILHFNHSGICELHCP-ALVFNFTVSWFLRVPKVSASHLERYTTFGASCVTAC 305 |
| Db | 238 | PDQTDCAFHRFNDGACVRCRCPQLVYNNKLTf-----QLBPNPHTKYQYGGVCVASC 290 |

| | | |
|----|------|---|
| Qy | 306 | PYNLSTDVGSCTLVCPHLHNOEVTAEDGTQRCCKSKPCARVCYVGLGMEHLREVRVTS 365 |
| Db | 291 | PHNFV-VDTSCVRACPPDKMEVD-KNGLKMKCEPCGGLCPKACEGTGSG--SRFQTVDS 346 |
| Qy | 366 | NIQFACGCKIFGSLAFLESFDCGDPASNTAPLOPELOVFTLEBEITGLYLIASWPDSL 425 |
| Db | 347 | NIDGFVNCTKILGNLDFLITGLNGDPWHKI PALDPEKLNLFRTVREITGLYLIQSWPPHM 406 |
| Qy | 426 | PDLVSFQNLQVIRGIRLHNGAYS-LTLOGLIGISMLGRSLRSLRELGLGLIHHNTHLCFVH 484 |
| Db | 407 | HNFSVFSNLTITIGRSLYNRGFSLLIMKNLVNVTSLGFRSLKEISAGRIYISANROLCYHH 466 |
| Qy | 485 | TPWDQLFRNPQALLHTA-NRDECECVGEGACHOLCARGHCWGPGPTQCVCNCSOFLRG 543 |
| Db | 467 | SLNMTKVLRGTEERLDIKHNRPRRDCVAEGKVCPLCSGGCGWPGQCCLSCRNYSRG 526 |
| Qy | 544 | QECVEECRVLQGLPREYVYARHCLPCHPECOQNGSVTCFGEADOCVACAHYKOPPFVC 603 |
| Db | 527 | GVCVTHCNFLNGEPREFAEAECSFCHPECOQMEGTATCNGSGDCAQCAHFRDGPCHVC 586 |
| Qy | 604 | ARCPGKVPDLSYMPIWKFPDEEGACQCPINCTHSCVLDLDDKGCPEAQR-----SPLTS 659 |
| Db | 587 | SSCPHGVLG--AKGPIYKYPDVQNECRPCHENCTQCGCKPELODCLGQTLVLIGKTHLTM 644 |
| Qy | 660 | IVSAVVGILLVVLGVVFGILIKRRQOKIR-KYTMRRLLQETELVEPLTPSGAMPNQAM 718 |
| Db | 645 | ALTVIAG--LVVIFMMLGGTFLYWRGRIQNKAMRRYLERGESIEPLDPS-EKANKVLA 701 |
| Qy | 719 | RILKETELRKVVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAY 778 |
| Db | 702 | RIFKETELRKVLGSGVFGTVHKGWVPEGESIKIPVCIKVIDKSGRQFQAVTDHML 761 |
| Qy | 779 | VMAVGVPYSRLIGICLTSTVOLVTOLMPYGCLLDHVRENRRGLSGDOLLNMCQIANG 838 |
| Db | 762 | AIGSLDHAHIVRLGLCPGSSQLQVTLPLGSLLDHVRQHGALGPOLLNMGVQIANG 821 |
| Qy | 839 | MSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIMMALE 898 |
| Db | 822 | MYLLEHGMVHRNLAAARNVLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKMALE 881 |
| Qy | 899 | SILRRRTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPOLLEKGERLPQPPCTIDVY 958 |
| Db | 882 | SIHFQKTHQSDVMSYGVYVWELMTFGAEPYAGLRLAEVDPDLLEKGERLAQPICTIDVY 941 |
| Qy | 959 | MIWVKWMIDSECPREFELVSEFMRMARDPQFVVIQNEIDLGA---SPLDSTFYRSL 1015 |
| Db | 942 | MMVVKWMIDENIRPTFKELANEFTMRMARDPRLVIVIKRES-GFOIACPEPHGLTNKKL 1000 |
| Qy | 1016 | EDDMGDLVDABEYLVPOQGFCCPDPAPGAGGMVHRRHRSSTRSGGDLTLGLEP-SEE 1074 |
| Db | 1001 | EEVELEPELDLDLEAED-----NLATTLGSALSIPVGTLPNP 1041 |
| Qy | 1075 | EAPRSLAPSEGAGSDVFDGDLGMAAKGLQSLPTH-PSPLQRYSEDPVPLP-----1127 |
| Db | 1042 | RGQSLSLPSGGY-MPMNQNLGSCQESAVSGSSERCPRVSLH-----PMRPGCLAS 1094 |
| Qy | 1128 | SETDGYVA-----PLTCSPOPE---YVNPQDVRPQPPSPREG-1162 |
| Db | 1095 | ESSEGHTVGEAELOEKVMCRSRSRSPRGRDSAYHSQRHSLLTPTVPLSPGLLEE 1154 |
| Qy | 1163 | -----LPAARPAAGATLERAKTLSP-GKNGVY-----KOVFAFGAVENYEYITPQCGA 1209 |
| Db | 1155 | DVNGYVMPDTHLKGTPSSREGTSLSSVLGSLGTGTEEBEDED-----EVEYMNRRRH 1206 |
| Qy | 1210 | APQHPPPAPSPAFDNLVYWD-----QDPPPERGAPSTFKGTPTAENPEYL 1255 |
| Db | 1207 | SP-PHPPRPSLBELGYEYMDVGSLSASLGSTQSCPLHPVPIIMPAGTTTDEDEYEM 1263 |

RESULT 10
ID_ERB3_RAT STANDARD; PRT; 1339 AA.
AC Q62799; Q62955;

15-DEC-1998 (Rel. 37, Created)
 15-JUN-2002 (Rel. 41, Last sequence update)
 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbb-3 precursor (EC 2.7.1.112)
 (c-erbB3).
 GN ERBB3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96096535; PubMed=8522190;
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 recombinant protein.";
 RL Gene 165:279-284(1995).
 RN [2]
 RP REVISIONS TO 85: 513 AND 565.
 RA Hellyer N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Fehner P.W., Kim S.S., Corbett J.A.;
 RT ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTA.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.
 CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 (POTENTIAL).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; U29339; AAC28498.2; -
 EMBL; U52530; AAC53050.1; -
 HSP; P11362; IFGK.
 InterPro: IPR000494; EGFR_L domain.
 InterPro: IPR000719; Euk pkinae.
 InterPro: IPR002174; Furin-like.
 InterPro: IPR001245; Tyr pkinae.
 Pfam: PF00069; pkinae; 1.
 Pfam: PF00757; Furin-like; 1.
 Pfam: PF01030; Recep_L domain; 2.
 PRINTS; PR00109; TYRKINASE.
 ProDom; PD000001; Euk pkinae; 1.
 SMART; SM00261; FU; 5
 SMART; SM00219; TyrK; 1.
 PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
 PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
 FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).

TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT BIND 713 721
 FT BIND 740 740
 FT ACT SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235
 FT DISULFID 231 243
 FT DISULFID 246 255
 FT DISULFID 259 286
 FT DISULFID 290 301
 FT DISULFID 305 320
 FT DISULFID 323 327
 FT DISULFID 500 509
 FT DISULFID 504 517
 FT DISULFID 520 529
 FT DISULFID 533 549
 FT DISULFID 556 573
 FT DISULFID 576 585
 FT DISULFID 589 610
 FT DISULFID 613 621
 FT DISULFID 617 629
 FT CARBOHYD 126 126
 FT CARBOHYD 250 250
 FT CARBOHYD 353 353
 FT CARBOHYD 408 408
 FT CARBOHYD 414 414
 FT CARBOHYD 437 437
 FT CARBOHYD 469 469
 FT CARBOHYD 522 522
 FT CARBOHYD 566 566
 FT CARBOHYD 616 616
 FT CONFLICT 1028 1028
 SQ SEQUENCE 1339 AA; 147545 MW; 0AASF2402BBFDF1E CRC64;
 Query Match 34.1%; Score 2329.5; DB 1; Length 1339;
 Best Local Similarity 40.5%; Pred. No. 7.3e-118;
 Matches 523; Conservative 173; Mismatches 431; Indels 163; Gaps 36;
 Qy 3 LAALRWGLLLALLPPGAA---STQVCTGTDMLRLPASPETHLDMLRHLYGCGQVOGN 59
 Db 7 LQVLC-----FLLSLARGSEMNSQAVCPGTLNGLSVTGDADNQYQTLKLYEKCEVVMGN 62
 Qy 60 LELTYLPTNASLSFLQDIOEQGVYVLIAHNOVROVPLQRLRIVRGTLQDFEDNYALAVLDN 119
 Db 63 LEIVLTGHNADLSFLQIREVTGYVLVANNEFSLPLNLRVVRGTQYVDGKFAIFVM-- 120
 Qy 120 GDPLNNTTPVTGASPGGLURELQLRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHNKOL 179
 Db 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLCMDTIDWRDIVRVR--- 170
 Qy 180 ALTLIDTNRSPACHPCSPKSGRCWGESSEDCQSLTFTVCAGGC-ARCKGPLPTDCCHE 238
 Db 171 GAELIVKXNGANGCPPEVCEKCG-KCWGPGDPDQQLTKTKTICAPQCNGRCFCGPNPQCCHD 229
 Qy 239 QAAAGCTGPKHSDCLACLHFNHSGICELHCP-ALVFNNTVSVFWRVPRKVSASHLERYTF 297
 Db 230 ECAGCGSPQDTDCFACRRFNDSGACVRCRPELVYKLTFT-----QLENPHTKYQY 282
 Qy 298 GASCVTACPNYSLTDVSGCTLVCPPLHNQEVTAEDGTQRCCKSKPCARVCYGLQWHLR 357
 Db 283 GGVCVASCPHNFV-VDQTFVCVRCAPPDKMEVD-KHGLKMCPCGGLCPKACGTCGSG--S 338
 Qy 358 EVRAVTSANIOEFACCKKIFGSLAFLPESFGDPPASNTAPLOPEQLQVPELEEITGYLY 417
 Db 339 RYQTVSSNIDGTFVNCYKILGNLDFLTGLNVDPMHKIPALDPEKLNFRVREITGYLN 398
 Qy 418 ISAWPDSLPDLVSFQNLQVIRGRILHNGAYS-LTLQGLGISWLGRLSRLSGSLALITHH 476

Db 399 IQSWPPHNFVSFNLTTIGRSLYNRFSLIMKLNINVTSLGFRSLKEISAGRVYISA 458
Qy 477 NTHLCFVHTVPDQFLRNPHQALLHTA-NRDECECVGEGALACHQLCARGHGWGPGTQCV 535
Db 459 NOQLCYHSLNWLTRLLRPSSEERLDIKYDRPLGECLAEGKVDPLCSSGGCGWPGQGL 518
Qy 536 NCSOFLRQOEVEECRVLOGLPREYVNRHCLPCHPECPONGSVTCFPGADQVACAH 595
Db 519 SCRNYSREGVCTHCFNLQGPREFVHEAQCFSCHPECLPMEGTSTCNGSGSDACARCAH 578
Qy 596 YKDPPEFCVARGSPGVKPDLSYMPKTFPDEGACOPCPINCHSC--VLDLDDKCPAQR 653
Db 579 FRDGHCVNSPHGILG--AGPIYKIDPAQNECRPHCHENTCQNGPELQDCLQAGAEVL 636
Qy 654 ASPLTSIVSAVGVLLVVLGVFGILIKRQOKIR-KYTMRRLLQETELVEPLTPSGAM 712
Db 637 MSKPHLVIAVTVG--LAVILMILGSGFLYWRGRRIQKRAMRRLYERGESIEPLDPS-EK 693
Qy 713 PNOQMRLKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 772
Db 694 ANKVLARIFKETELRKVLKVLGSGVGTGVIKWIPEGESIKIPVCIKVIEDKSGRSQFA 753
Qy 773 ILDEAYVMAGVSPVSRLLIGLICTSTVQLVTLQMLPYCCLLDVHRENRGLSGQDLNWC 832
Db 754 VTDHMLAVGSLDHAHIVLLGLCPGSSQLVTVPLGSLLDHVKQHRRTGLPQLLLNWG 813
Qy 833 MQIAKMSYLEVDRLVLRDLAARNVLKSPNHRKITDFGLARLLDIDETEHADGKVP 892
Db 814 VOIAKGMVYLEHSMVHRDLARNVLMKSPSOVQVADGVADLLPDDKQLLHSEAKTP 873
Qy 893 KMALESILRRFTHQSDVMSYGVTVWELMTFTGAKPYDGIIPAREIPDLLEKGERLPQPI 952
Db 874 KMALESIFHGKYTHQSDVMSYGVTVWELMTFGAEPYAGLRLAIPDLLEKGERLAQPI 933
Qy 953 CTIDVVMKWKMIIDSECRPRFELYSERFMRDQRFVVIQNEIDLGPASPLDSTFYR 1012
Db 934 CTIDVVMKWKMIIDENIRTFKELANEFTFMRDQRFVVIQNEIDLGPASPLDSTFYR 990
Qy 1013 SLEDDDMGDLVDAEYLVPOQGFCDPAPGAGVWVHRRSSRSGGDLTLGLEPS 1072
Db 991 SVLTTKEL---QEALEPEL-----DLDLDELE 1016
Qy 1073 EE-----EAPRPLAPSEG-----AGSDVFDGLGMAKGL 1104
Db 1017 EELGATSLGALSALPTGTLRPGSQSLSPSSGYMPMNOSSGEACLDASVILGREGQFS 1076
Qy 1105 QSLPHTDPSLQRYSEDPTVPLPSETGVY---APL-----TC-----SPOPE---- 1144
Db 1077 RPISLH-PIPRGR-----PASESEGHVTGSAELQEVSVCRSRSRSPRPGDSA 1128
Qy 1145 YVNPQDVRPOPSPREGP-----LPAARPAGATLERAKTSLP-GKNGVY----- 1187
Db 1129 YHSQRHLLTPVTPLSPLGLEEDGNGVMPDTHLRCGASSSREGTSSVGLSSVLGTTEE 1188
Qy 1188 -KDVFAFGGAVENPEYLTPOGGAAPQHPH 1216
Db 1189 DED-----EYEVNMRKRGSP-PRPP 1209

RESULT 11
EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (torpedo protein) (Drosophila relative of ERBB)
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.
RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; Tissue=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J.J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Baldwin D.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,


```
Db 423 IDGNIRLDQTFSGFDQVYANTMGPRYIPDPERREVFSTVKEITGYLNIEGTHPQPRN 482
Qy 428 LSVFQNLQVIRGLHNGAY-SLTQGLGIGISWGLRLSRLGSLALIHNNHLCFVHTV 486
Db 483 LSYFRNLETHGRQLMESFAALAIVKSSLYSLEMRNLKQISSGWSVVIQHRDLCCVSN 542
Qy 487 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCHGPGPTQCVCNSQFLRGQEC 546
Db 543 RWPALQKEPEQKVMWNLRLADLCEKNGTI CSDQCNEDCGWAGAGDQCLTKCNFNFGTC 602
Qy 547 VEECRVQLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPPEFCVARC 606
Db 603 IADCGYISNAYK--FDNPTCKICHPECR-----TCNAGADHCOECVVRDQGHCVSEC 654
Qy 607 P-----SGVK-----PD 614
Db 655 PKNYNDRGVCRECHATCDGCTGPKDTIGIGACTTCNLAIINNDATVRCLLKDDKCPD- 713
Qy 615 SYMPIWKF--PDEGACOP-----CPI-----NCTH----- 638
Db 714 GY---FWEYVHPQEOGSLPLAGRAVCRKCHPLCELCTNYGHEQVCSKTHYKREOCET 771
Qy 639 -----SC-----VDLDDKG----- 647
Db 772 ECPADHYTDEQRECFQHPPECNGCTGPGADDCKSCRNFKLFDANETGPNVSTWFNCTS 831
Qy 648 -CPAEQR-----ASPLTS-----IVSAVWGLLVVLGVVFGI 679
Db 832 KCPLMRHVNYQYTAIGPYCAASPRSSKITANLDVNMIFIITGAVLVPTICILCV--T 889
Qy 680 LIKROOKIRKYT--MRLLOTELVELPTSGAMPNQAOHRIKTELRLKVKVLGSGAF 737
Db 890 YICQKQKAKKETVMYTMALSGCDESEPLRPSNTGANLCKLRIVKDAELRKGGLGMGAF 949
Qy 738 GTVYKGIWIPDGENVKIPVAIKLRENTSPKANKEILDEAVVWAGVSPYVSRLLIGLT 797
Db 950 GRVYKGVWPEGENVKIPVAIKELKSTGAESSEFELEAVIMASEEHVNLKLLAVCMS 1009
Qy 798 STVOLVTQMPYGCLLDHVRENRLGSLQDILLNMCQIAKMSYLEDLVRLVHRLAARNV 857
Db 1010 SOMMLITQLMPLGLCLDYVRNRNDKIGSKALLNWSQIAKMSYLEEKRLVHRLAARNV 1069
Qy 858 LVKSPNHVKITDFGLARLLIDETEHYADGCKVPIKMALESILRRPETHOSDWSVGV 917
Db 1070 LVQTPSLVKITDFGLAKLLSSDSNEYKAAAGKMPKMLALECIRNRVFTSKSDVWAFGV 1129
Qy 918 VWELMTGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVVKMIDSECRPRFRE 977
Db 1130 IWELMTFGORPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTILLSCHWLDAAVRPTKQ 1189
Qy 978 LVSEFSRMDPQRFVYQNEIDLQ--PASPLDSTFYRSLLEDD--DMGDLVDAAEYLVP 1032
Db 1190 LTTVFASFARDPGRYLAIPGDKFRLPA-----YTSQDEKDLIRKLAPTPTDSEAIK 1242
Qy 1033 QOQFFCDDPAPGAGVMVHRRSSRSTSGGDLTLGLEPSEEP-----RSLPASEG 1086
Db 1243 PDDYLQKAAPGPS-----HRTDCT-----DEMPKLNRYCKDPSNKNSS 1281
Qy 1087 AGSDVDFG---DLGMAAGKGLQSLTHDPSPQLRYSDEPTVPLPSETGDVYAPLTCSPQ 1143
Db 1282 TGDDESSAREVGVGNLR-----LDLPVDEDDYLMPTCQGP 1319
Qy 1144 EYVNPQDVRPQPPSPREGPIPAARPAGATLERAKTSLPGKNGVVKQVFAFGAVENPEYL 1203
Db 1320 NNNNNNN-----NPNQNNMAAVGAAGYM-----DLIGVPVSDNPEYL 1358
Qy 1204 ----TPQGAAPQPH-----PPPAFSP-AFNLVYWD 1230
Db 1359 LNAQTLGVGSEPIPTQITIGIPVMGGPGTMEVKVPMGSEPTSSDHEYND 1408
```

RESULT 12

```
ERBB_ALV
ID ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -|- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFVLV.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
DR Glycoprotein; Phosphorylation.
DR DOMAIN 132 399 PROTEIN KINASE.
DR NP_BIND 138 146 ATP (BY SIMILARITY).
DR BINDING 165 165 ATP (BY SIMILARITY).
DR ACT_SITE 257 257 BY SIMILARITY.
DR ACT_SITE 257 257 BY SIMILARITY.
DR SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
Query Match 25.6%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 4.5e-87;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;
```

```
Qy 593 CAHYKDPFPFCVACPSGVKPDLSYMPITWKPDEGACQPCPINCTHSCVDLDDGCPAEQ 652
Db 3 CAHFIDGPHCVKACPCAGVGLGENDTL-VMKYADANAVCOLCHPNCRTGCKGFGEGCP--- 58
Qy 653 RASPLTSIVSAV-VGILLVVLGVVFGILIKRQOKIRKYTMRLLOSTELVELPTPSGA 711
Db 59 NGSKTFSIAGVVGGLLCLVVGIGIGLYLRRR-HIVKRTLRLQLRELVELPTPSGE 117
Qy 712 MPNOAQMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKLRENTSPKANK 771
```

```
118 APNOAHLRLKETETPKVKVLGSGAFGTYYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
QY 772 EILDEAYVMAGVSGPYVSRLLGICLTSTVLTQMLPYGCLLDHVRNRLGSDLLNW 831
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVLTQMLPYGCLLDYIREHKDNTGSOYLLNW 237
QY 832 CWOIAGMSYLEDVRLVRLDAAARNVLKSPNHVKITDFGLARLLDIDETEYHAGGKVP 891
Db 238 CVOIAGMNYLEERLVRDLAARNVLKTPQHVITDFGLAKLLGADKEKYEHAEGKVP 297
QY 892 IKWMALESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIPIAREIPDILLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVVWELMTFGSKPYDGIPIAREIPDILLEKGERLPQPP 357
QY 952 ICTIDVYIMVKCMWIDSECRPRFRELVSERMDARQRFVVIQ-NEDLGASPLDSTF 1010
Db 358 ICTIDVYIMVKCMWIDSECRPRFRELVSERMDARQRFVVIQ-NEDLGASPLDSTF 417
QY 1011 YRSLEDDMDGLVDAAEYLVPOQGFPCDPAPGAGGMVHRRHSSTRSGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1071 PSEEAAPSPL-----APSEGAGSDVFDGLGMGAAGLQSLTPHDPSPQRYSEDPVTP 1125
Db 450 -----SRTPLLSLSATSNSTNCID-----RNGQGHVREDSPVQRYSDPTCN 495
QY 1126 LPSET--DGYVAPLTCSPQEVVQPDVROPSPREGPLPAARAGATLERAKTLPCK 1183
Db 496 FLEESIDGFL-----PAPEYVNO--LMPKPKS-----TAMVQ 526
QY 1184 NGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQHPPPAFSPADNLY 1228
Db 527 NQIYNNISITAIKSLPMDSRYQNSHSTAVDNEYL-----NTNQSPLAKTVPFESSPY 578
QY 1229 WDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLOVP 1260
Db 579 WIQSGNHQINDPNPDYQDFLNETKPNGLLKVPAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaisa M., Biserre G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
```

```
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A06644; TVYUH.
DR HSP; F11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACET_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
FT SEQUENCE 604 AA; 67633 MW; 76EBCDD06745D609 CRC64;
Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 1.4e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;
QY 593 CAHYKDPFPCVACPSGVKPDLSYPIWKFPDEACQPCINCHTSCVDLDDKGPAPQ 652
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VRKYADANAVCQLCHPNCCTRGCKGPGLEGCP--- 58
QY 653 RASPLTSIVSAV--GILLVWVGVVGLIKRQOKIRKYTWRLLOETELVEPLTPSGA 711
Db 59 NGSKTPSTAAGVVGGLCLVVGIGLYLRR--HIVKRTLRLLQERLVEPLTPSGE 117
QY 712 MPNOAMRLKETELRKVKVLGSGAFGTYYKGLWIPEGEKVKIPVAIKELREATSPKANK 771
Db 118 APNOAHLRLKETETPKVKVLGSGAFGTYYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
QY 772 EILDEAYVMAGVSGPYVSRLLGICLTSTVLTQMLPYGCLLDHVRNRLGSDLLNW 831
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVLTQMLPYGCLLDYIREHKDNTGSOYLLNW 237
QY 832 CWOIAGMSYLEDVRLVRLDAAARNVLKSPNHVKITDFGLARLLDIDETEYHAGGKVP 891
Db 238 CVOIAGMNYLEERLVRDLAARNVLKTPQHVITDFGLAKLLGADKEKYEHAEGKVP 297
QY 892 IKWMALESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIPIAREIPDILLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVVWELMTFGSKPYDGIPIAREIPDILLEKGERLPQPP 357
QY 952 ICTIDVYIMVKCMWIDSECRPRFRELVSERMDARQRFVVIQ-NEDLGASPLDSTF 1010
Db 358 ICTIDVYIMVKCMWIDSECRPRFRELVSERMDARQRFVVIQ-NEDLGASPLDSTF 417
QY 1011 YRSLEDDMDGLVDAAEYLVPOQGFPCDPAPGAGGMVHRRHSSTRSGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1071 PSEEAAPSPL-----APSEGAGSDVFDGLGMGAAGLQSLTPHDPSPQRYSEDPVTP 1125
Db 450 -----SRTPLLSLSATSNSTNCID-----RNGQGHVREDSPVQRYSDPTCN 495
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds
(without alignments)

5347.444 Million cell updates/sec

Title: SEQ4-250-264-14

Perfect score: 6839

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGIDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_xvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 6191 | 90.5 | 1259 | 6 | O18735 canis famil |
| 2 | 3119 | 45.6 | 1209 | 11 | Q9QX70 |
| 3 | 3090 | 45.2 | 1210 | 11 | Q9EP98 |
| 4 | 2724 | 39.8 | 1165 | 13 | Q9YH40 |
| 5 | 2704.5 | 39.5 | 1137 | 13 | Q9W6F6 |
| 6 | 2284 | 33.4 | 1328 | 13 | Q9W6F6 |
| 7 | 2011.5 | 29.4 | 1433 | 5 | Q9B1H9 |
| 8 | 1770.5 | 25.9 | 419 | 4 | Q9UK79 |
| 9 | 1739 | 25.4 | 367 | 11 | Q8R2X1 |
| 10 | 1720 | 25.1 | 729 | 15 | Q86712 |
| 11 | 1718 | 25.1 | 567 | 15 | Q86714 |
| 12 | 1697.5 | 24.8 | 412 | 4 | Q8WYV0 |
| 13 | 1653.5 | 24.2 | 962 | 15 | O64895 |
| 14 | 1645 | 24.1 | 545 | 15 | O85468 |
| 15 | 1481.5 | 21.7 | 655 | 11 | Q9WVF5 |
| 16 | 1465.5 | 21.4 | 643 | 11 | Q9ERV6 |

| | | | | | |
|----|--------|------|------|----|--------|
| 17 | 1240 | 18.1 | 1193 | 5 | Q9Y1X8 |
| 18 | 1157.5 | 16.9 | 1368 | 5 | Q23821 |
| 19 | 1146 | 16.8 | 1717 | 5 | Q26566 |
| 20 | 1103 | 16.1 | 527 | 13 | Q90836 |
| 21 | 976.5 | 14.3 | 478 | 11 | Q9ESE0 |
| 22 | 969.5 | 14.2 | 599 | 13 | Q9PSH2 |
| 23 | 906 | 13.2 | 165 | 4 | Q14256 |
| 24 | 887 | 13.0 | 176 | 11 | Q923V5 |
| 25 | 806.5 | 11.8 | 346 | 13 | P11776 |
| 26 | 778 | 11.4 | 435 | 5 | Q8SZW1 |
| 27 | 754.5 | 11.0 | 311 | 13 | Q99162 |
| 28 | 745.5 | 10.9 | 1362 | 13 | Q9PV24 |
| 29 | 725 | 10.6 | 1671 | 5 | Q9NJV5 |
| 30 | 713.5 | 10.4 | 1368 | 13 | Q8UW85 |
| 31 | 702 | 10.3 | 1418 | 13 | Q93457 |
| 32 | 694 | 10.1 | 331 | 4 | Q9BUD7 |
| 33 | 686.5 | 10.0 | 1369 | 13 | Q8UW86 |
| 34 | 678 | 9.9 | 1358 | 13 | Q73798 |
| 35 | 656.5 | 9.6 | 1472 | 5 | Q9U5A8 |
| 36 | 651 | 9.5 | 1412 | 13 | Q8UW84 |
| 37 | 647 | 9.5 | 1245 | 13 | Q9YGH8 |
| 38 | 644.5 | 9.4 | 1418 | 13 | Q8UW83 |
| 39 | 635 | 9.3 | 1371 | 11 | Q9QW44 |
| 40 | 619 | 9.1 | 149 | 6 | Q9BG66 |
| 41 | 610.5 | 8.9 | 2144 | 5 | Q9VD94 |
| 42 | 589 | 8.6 | 987 | 11 | Q91YM0 |
| 43 | 587.5 | 8.6 | 1036 | 4 | Q07912 |
| 44 | 587.5 | 8.6 | 1055 | 11 | O54967 |
| 45 | 586.5 | 8.6 | 1091 | 4 | Q9UMQ4 |

ALIGNMENTS

RESULT 1

| | | | |
|--------|---|---|---------------|
| Q18735 | O18735 | PRELIMINARY; | PRT; 1259 AA. |
| AC | O18735; | | |
| DT | 01-JAN-1998 | (TremBLrel. 05, Created) | |
| DT | 01-JAN-1998 | (TremBLrel. 05, Last sequence update) | |
| DT | 01-JUN-2002 | (TremBLrel. 21, Last annotation update) | |
| DE | ErB-2. | | |
| OS | Canis familiaris (Dog). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis. | | |
| OX | NCBI_taxID=9615; | | |
| RN | [1]_taxID=9615; | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | Yokota H.; | | |
| RT | "cDNA cloning of erbB-2 from canine mammary gland."; | | |
| RL | Submitted (OCT-1997) to the EMBL/GenBank/DBAJ databases. | | |
| DR | EMBL; AB008451; BAA23127.1; -. | | |
| DR | HSSP; P11362; IFGK. | | |
| DR | InterPro; IPR002048; EF-hand. | | |
| DR | InterPro; IPR000494; EGFR_L domain. | | |
| DR | InterPro; IPR000719; Euk_Pkinase. | | |
| DR | InterPro; IPR002174; Furin-like. | | |
| DR | InterPro; IPR001245; Tyr_pkinase. | | |
| DR | InterPro; IPR004019; YLP_motif. | | |
| DR | Pfam; PF00757; Furin-like; 1. | | |
| DR | Pfam; PF00069; pkinase; 1. | | |
| DR | Pfam; PF01030; Recep_L_domain; 2. | | |
| DR | Pfam; PF02757; YLP_2_domain; 1. | | |
| DR | ProDom; PD000001; Euk_pkinase; 1. | | |
| DR | SMART; SM00261; FU; 3. | | |
| DR | SMART; SM00219; TyrKc; 1. | | |
| DR | PROSITE; PS00018; EF_HAND; UNKNOWN 1. | | |
| DR | PROSITE; PS00107; PROTEIN_KINASE_ATP; 1. | | |
| DR | PROSITE; PS50011; PROTEIN_KINASE_DOM; 1. | | |
| DR | PROSITE; PS00109; PROTEIN_KINASE_TYR; 1. | | |
| KW | ATP-binding; Transferase; Tyrosine-protein kinase. | | |
| SQ | SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64; | | |

| | | | | | |
|-----------------------|------|---------------------|--------------------------|--------------------------------|--------------|
| Query Match | | 90.5%; | Score 6191; | DB 6; | Length 1259; |
| Best Local Similarity | | 90.6%; | Pred. No. 0; | | |
| Matches 1147; | | Conservative 41; | Mismatches 66; | Indels 12; | Gaps 3; |
| QY | 1 | MELALACRWGLLLALLP | PGAASTQVCTGDMKRLRPAS | PETHLDMRLHLYQCGVQVQNL | 60 |
| DB | 1 | MELAAWCRWGLLLALLP | PSGAAGTQVCTGDMKRLRPAS | PETHLDMRLHLYQCGVQVQNL | 60 |
| QY | 61 | ELTYLPTNASLFLQDI | QIEVQGVLIHNOVQVPLQRLRI | VRGTQFDPEDNALAVLDNG | 120 |
| DB | 61 | ELTYLPANASLFLQDI | QIEVQGVLIHNSQVRIPLQRLRI | VRGTQFDPEDNALAVLDNG | 120 |
| QY | 121 | DFLNNTPVTGASPGGLREL | QLRSLEILKGGVLIQRNPOLCYQD | TILWKDIFHKNNQLA | 180 |
| DB | 121 | DPLEGGIPAPGAAQGLREL | QLRSLEILKGGVLIQRSPOLCHQD | TILWKDVFHKNNQLA | 180 |
| QY | 181 | LTLIDTNRSRACHPCSP | CMKGRWCESSEDCOSLTRVCAG | GCARCKGPLTDCHEQC | 240 |
| DB | 181 | LTLIDTNRSAKPPCSP | PACKDAHCWASSGDCQSLTRVC | AGGCARCKGPQPTDCHEQC | 240 |
| QY | 241 | AAGCTGPKHSDCLAC | LHFNHSGTICELHCPALVFNNPT | VSFWLRVPKVSASHLERYTFGAS | 300 |
| DB | 241 | AAGCTGPKHSDCLAC | LHFNHSGTICELHCPALVTYNTD | TFSMPNPF-----GRYTFGAS | 294 |
| QY | 301 | CVTACPYNYLSTDVGS | CTLVCPPLHNQEVTAEDGTQRC | CEKSKPCARVCYVGLGMEHLREVR | 360 |
| DB | 295 | CVTSCPYNYLSTDVGS | CTLVCPPLNNQEVTAEDGTQRC | CEKSKPCARVCYVGLGMEHLREVR | 354 |
| QY | 361 | AVTSANIQEFACKKIF | SGSLAFELPESFDGDPASNTAP | LOPEOLQVPETLEEITGYLYISA | 420 |
| DB | 355 | AVTSANIQEFACKKIF | SGSLAFELPESFDGDPASNTAP | LOPEQLRFEALEEITGYLYISA | 414 |
| QY | 421 | WPDLSLPDLSVFQNL | QVIRGIRILHNGAYSILTLOGL | ISWLGSLRSLGSLALIHNTHL | 480 |
| DB | 415 | WPDLSLPNLSVFQNL | VRIRGVLRHDCAYSILTLOGL | ISWLGSLRSLGSLALIHNRAL | 474 |
| QY | 481 | CFVHTVPMDLPNPHQ | ALLHTANRDECEVGEGLACHOL | CARGHGWGPGTQCVCNCSQF | 540 |
| DB | 475 | CFVHTVPMDLPNPHQ | ALLHSANRPEECVGEGLACYP | -CAHGHGWGPGTQCVCNCSQF | 533 |
| QY | 541 | LRGQCEVECRVLQGL | PREYVNVARUCLPCHPCOPONG | SVTCFGEADQVCACAHYKDDP | 600 |
| DB | 534 | LRGQCEVECRVLQGL | PREYVNDYRCLPCHSCQCPNG | SVTCFGEADQVCACAHYKDDP | 593 |
| QY | 601 | FCVARCPGVKPDLS | YMPYWKFPDEBAGCQPCIN | CTHSCVDLDDKGCPCAEORASPLTSI | 660 |
| DB | 594 | FCVARCPGVKPDLS | YMPYWKFADEBGTQCP | CINCTHSCALDEKGCPCAEORASPVTSI | 653 |
| QY | 661 | VSAAVGIILVVVLG | VVVGILIKRQOKIRKYTMRL | LLOTELVEPLTPSGAMPNQAQMRI | 720 |
| DB | 654 | IAAVVGIILVVVLG | VVVGILIKRRRQIRKYTMRL | LLOTELVEPLTPSGAMPNQAQMRI | 713 |
| QY | 721 | LKETELRKVKVLG | SGAFVYVGIWIPDGENVKI | PVAIKVLRNTSPKANKIILDEAYVM | 780 |
| DB | 714 | LKETELRKVKVLG | SGAFVYVGIWIPDGENVKI | PVAIKVLRNTSPKANKIILDEAYVM | 773 |
| QY | 781 | AGVGSPPYVSRLLG | ICLTSTVQLVTLQMPYGC | LLDHHVRENRLGSGODLLNMCQIAKMS | 840 |
| DB | 774 | AGVGSPPYVSRLLG | ICLTSTVQLVTLQMPYGC | LLDHHVREHRLGSGODLLNMCVQIAKMS | 833 |
| QY | 841 | YLEDVRLVHRDLA | ARNVLKSNHVKITDGLAR | LDDIDETEHADGGKVPVKKWMALES | 900 |
| DB | 834 | YLEDVRLVHRDLA | ARNVLKSNHVKITDGLAR | LDDIDETEHADGGKVPVKKWMALES | 893 |
| QY | 901 | LRRRTTHOSDWS | YGVYVWELMTFGAKPYDGI | PAIREIPDLLEKGERLPQPPCTTDVMI | 960 |
| DB | 894 | PPRRTTHOSDWS | YGVYVWELMTFGAKPYDGI | PAIREIPDLLEKGERLPQPPCTTDVMI | 953 |
| QY | 961 | MVKCMWIDSECR | PREFRELVSERMRARDPQ | REWIQNEDLGPASPLDSTFFRSILLEDDDM | 1020 |
| DB | 954 | MVKCMWIDSECR | PREFRELVAESMRARDPQ | REWIQNEDLGPASPLDSTFFRSILLEDDDM | 1013 |
| QY | 1021 | GDLVDAEYLV | VPQGGFFCPDPAPAGG | WVHHRRSSSTRSGGDLTLGLEPSEEPKSP | 1080 |

| | | | | | |
|-----------------------|--|---|---------------------|---|------|
| Db | 1014 | GDLVDAEYLV | VPQGGFFCPDPAGGTAHRR | SSSTRNGGELTLGLEPSEEPKSP | 1073 |
| Qy | 1081 | LAPSEGAGSDV | PDGLGMAAGKLSLPTHDS | PLQRYSEDPTVPLPSETDGYVAPLTCS | 1140 |
| Db | 1074 | LAPSEGAGSDV | PDGLGMAAGKLSLQSLPS | QDPSPLQRYSEDPTVPLPSETDGYVAPLTCS | 1133 |
| Qy | 1141 | POPEYVNP | PDVRPOPPSPREGPLPA | ARPAGATLER-----AKTLPFGKNGVVKDVFATGG | 1195 |
| Db | 1134 | POPEYVNP | PEVWPQPLALEGPLPS | RPAAGATLERKTLSPKTLSPKNTLSPFGKNGVVKDVFATGS | 1193 |
| Qy | 1196 | AVENPEYLT | TPQGGAAQPPHPPAF | SFADNLYYWDQDPPERCAPSTFKGTPTAENPEYL | 1255 |
| Db | 1194 | AVENPEYLA | PRGAAAPQPHPPAF | SFADNLYYWDQDPSERGSPTSTFEGTPTAENPEYL | 1253 |
| Qy | 1256 | GLDVPV | 1261 | | |
| Db | 1254 | GLDVPV | 1259 | | |
| RESULT 2 | | | | | |
| Q9QX70 | | | | | |
| ID | Q9QX70 | PRELIMINARY; | PRT; | 1209 AA. | |
| AC | Q9QX70; | | | | |
| DT | 01-MAY-2000 | (TRENBLrel. 13, Created) | | | |
| DT | 01-MAY-2000 | (TRENBLrel. 13, Last sequence update) | | | |
| DT | 01-JUN-2002 | (TRENBLrel. 21, Last annotation update) | | | |
| DE | Epidermal growth factor receptor. | | | | |
| GN | EGFR. | | | | |
| OS | Rattus norvegicus (Rat). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. | | | | |
| OX | NCBI_TaxID=10116; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=FISHER; TISSUE=LIVER; | | | | |
| RX | MEDLINE=90258888; PubMed=2342466; | | | | |
| RA | Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C., | | | | |
| RA | Earp H.S.; | | | | |
| RT | "A truncated, secreted form of the epidermal growth factor receptor is | | | | |
| RT | encoded by an alternatively spliced transcript in normal rat tissue."; | | | | |
| RL | Mol. Cell. Biol. 10:2973-2982(1990). | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=FISHER; TISSUE=LIVER; | | | | |
| RA | Petch L.A.; | | | | |
| RL | Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases. | | | | |
| RN | [3] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=FISHER; TISSUE=LIVER; | | | | |
| RA | Guttridge K., Dawson T.L., Earp H.S.; | | | | |
| RL | Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases. | | | | |
| DR | EMBL; M37394; AAF14008.1; - | | | | |
| DR | HSP; P11362; IFGK. | | | | |
| DR | InterPro; IPR000494; EGFR_L domain. | | | | |
| DR | InterPro; IPR000719; Euk_pkinase. | | | | |
| DR | InterPro; IPR002174; Furin-like. | | | | |
| DR | InterPro; IPR001245; Tyr_pkinase. | | | | |
| DR | Pfam; PF00757; Furin-like; 1. | | | | |
| DR | Pfam; PF00069; pkinase; 1. | | | | |
| DR | Pfam; PF01030; Recep_L domain; 2. | | | | |
| DR | PRINTS; PR00109; TYRKINASE. | | | | |
| DR | ProDom; PD000001; Euk_pkinase; 1. | | | | |
| DR | SMART; SM00261; FU; 3. | | | | |
| DR | SMART; SM00219; Tyrc; 1. | | | | |
| DR | PROSITE; PS00107; PROTEIN_KINASE_ATP; 1. | | | | |
| DR | PROSITE; PS00011; PROTEIN_KINASE_DOM; 1. | | | | |
| DR | PROSITE; PS00109; PROTEIN_KINASE_TYR; 1. | | | | |
| KW | ATP-binding; Receptor; Transferase; Tyrosine-protein kinase. | | | | |
| SQ | SEQUENCE 1209 AA; 134891 MW; 96FEE7F6CCLB7773 CRC64; | | | | |
| Query Match | | 45.6%; | | Score 3119; DB 11; Length 1209; | |
| Best Local Similarity | | 49.6%; | | Pred. No. 7.8e-227; | |

| | | |
|----|---|--|
| DR | PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1. | |
| DR | PROSITE; PS00107; PROTEIN KINASE ATP; 1. | |
| DR | PROSITE; PS00011; PROTEIN KINASE DOM; 1. | |
| DR | PROSITE; PS00109; PROTEIN KINASE_TYR; 1. | |
| KW | ATP-binding; Receptor; Transferase. | |
| SQ | SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64; | |
| | Query Match 45.2%; Score 3090; DB 11; Length 1210; | |
| | Best Local Similarity 49.1%; Pred. No. 1.2e-224; | |
| | Matches 628; Conservative 170; Mismatches 364; Indels 116; Gaps 24; | |
| QY | 11 LLALLPPGAA--STOVCTGTDMKRLPASPETHLDMLRLHLYGCGOVQGNLELTLPYN 68 | |
| DB | 14 LLTALCAAGALEEKVKCGTSCNRLTQLGTFEDHFLSLQRMNNYNCVWLNLEITVYQVN 73 | |
| QY | 69 ASLSFLDLOEVOGYVLIHAHNOVQVPLQRLRIVRGTOFLFEDNYALAVLDNGDPLNNTTP 128 | |
| DB | 74 YDLSFLTKTQEVAGYVLIALNTVERIPLENLQIRGNALYENTYALAILSN----- 124 | |
| QY | 129 VTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDI-----FKNNQLALTLI 184 | |
| DB | 125 -YGTNRTGLRELPMRLNLEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMDL--- 180 | |
| QY | 185 DTRNSRACHPCPMCKGSCWGESSEDCQSLRTVTCAGGCA-RCKGPLPTDCHEGCAAG 243 | |
| DB | 181 -QSHFSPCKPDCPSCPNCGSCWGGEBENCQKLTIIICAOQCSHRCRGRSPSDCHNQCAAG 239 | |
| QY | 244 CTGPKHSDCLACLFHNSGICELHCPALVFNFTVSFLMRVPKVSASHLERYTFGASCVT 303 | |
| DB | 240 CTGPRESCLVCKQFODEATCTDCTPPLMLNPTTY-----QMDVNPPEGKYSFGATCYK 293 | |
| QY | 304 ACPPNYLSTDVGSCTLVCLPHNQEVTAEDGTORCEKSKPCARVCYVGLGMEHLREVRVAT 363 | |
| DB | 294 KCPRYVVTVDHSCVRACGPDYEV-EEDGIRKCKDCGPKVCNGIGIGEFKDTLSIN 352 | |
| QY | 364 SANIOEFACKKIFGSLAFPLPSFDGDPASNTAPLOPEOLQVPETLEEITGYLYISAWPD 423 | |
| DB | 353 ATNIRKFKYCTAISGDHLPLVAFKGDSTFTPLDPRELEILKTVEITGFLLIQAWPD 412 | |
| QY | 424 SLPDLSVFONLOVIRGRIHLNAGYSLTLOGLSWGLSLRSLRSLGSLALIHNTHLCFV 483 | |
| DB | 413 NWTDLHAFENLEIRGTRKQHQFSLVAVGLNITSGLRSLKEISDGDVIIISGNRLCYA 472 | |
| QY | 484 HTVPMDQFRNPQHALLHTANRPEDECVGEGGLACHOLCARHGCGPGTQCVCNCSQFLRG 543 | |
| DB | 473 NTINNKLFGTNPQKTKIMNNAEKDCKAVNHVCNPLCSSEGCWGPEDPCVSCQNVSRG 532 | |
| QY | 544 QECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCV 603 | |
| DB | 533 RECVEKCNILEGEPREFVENSEICQHPCELPQAMNITCTGRGPDNCIOCAHYIDGPHCV 592 | |
| QY | 604 ARCPGSKPDLVMPITWKFPEDEGACQPCPINTHSCVDLDDKGCPEAQRASPLTSIVSA 663 | |
| DB | 593 KTCPAGINGENTL-VWKYADANNVCHLCHANCTYGCAGPGLGCEVWSPGKPIPSIATG 651 | |
| QY | 664 VVGILLVVLGVVFGI-LIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRLIK 722 | |
| DB | 652 IVGGLLFIW-VALGIGLFMRRRHVRKTRLLRLLQERELVEPLTPSGEAPNQAHRLIK 710 | |
| QY | 723 ETELKRVKVLGSGAGTYVKGIIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 782 | |
| DB | 711 ETEFKKIKVLGSGAGTYVKGIIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAS 770 | |
| QY | 783 VGSPPVYVRLGLCLTSVTOLVLTQMLPYGCLLDHVRNRRGLSQDILLNWCMOIAGMSYL 842 | |
| DB | 771 VDNPHVCRLLGLCLTSVTOLITQMLPYGCLLDYVREHKDNIGSYQLLNCWCQVQIAGMNYL 830 | |
| QY | 843 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKMWALESIIR 902 | |
| DB | 831 EDRLVHRDLAARNVLKTPQHVKITDFGLAKLGAEEKEYHAEGKVPKMWALESIILH 890 | |
| QY | 903 RRFTHQSDVWSYGVTVWELMTFGSKPYDGI PASDISSILEKGERLPQPPICTIDVYIMV 962 | |

| | | |
|----------|--|--|
| DB | 891 RIYTHQSDVWSYGVTVWELMTFGSKPYDGI PASDISSILEKGERLPQPPICTIDVYIMV 950 | |
| QY | 963 KCMWIDSECRPRRELVSFESRMARDPQFVVIQ-NEDLGPASPLDSTFYRSLLEDDMG 1021 | |
| DB | 951 KCMWIDADSRPKPRFRELILFESKWARDPQRYLVITQGDHRLPSPSTNSFYRALMDEDEME 1010 | |
| QY | 1022 DLVDAEYLVPOQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSPL 1081 | |
| DB | 1011 DVVDADEYLTPOQGF-----NSPST-----SRTPL 1036 | |
| QY | 1082 APSEGAGSDVFDGDLGMGAAGLQSLPTHDPSPQRYSEDPTVPLPSET--DGYVAPLTC 1139 | |
| DB | 1037 LSSLSATS----NSTVACINRNGSCRVEDAFQRYSSDPTGAVTEDNIDDAFL----- 1087 | |
| QY | 1140 SPOPEVNOQDVPYRPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVEN 1199 | |
| DB | 1088 -VPVEYVNO-SVPKRPAGSQVNPVYHNPQLHP-----APGRDLHYQN--PHSNAVGN 1135 | |
| QY | 1200 PBYL-TPQCGAAPQHPHPPAFSPAFONLYWQD-----DP-----PERGAPPST 1242 | |
| DB | 1136 PEYLNATQ-----PTCLSSGFNSPALWIKGSHQMSLDNPDYQDDFFPKETKPNGI 1186 | |
| QY | 1243 FKGTPTAENPEYLGLDVP 1260 | |
| DB | 1187 FKGTPTAENPEYLGLDVP 1203 | |
| RESULT 4 | | |
| QY | QYH40 PRELIMINARY; PRT; 1165 AA. | |
| ID | QYH40; | |
| AC | QYH40; | |
| DT | 01-MAY-1999 (TREMELrel. 10, Created) | |
| DT | 01-OCT-2000 (TREMELrel. 15, Last sequence update) | |
| DT | 01-JUN-2002 (TREMELrel. 21, Last annotation update) | |
| DE | Receptor tyrosine kinase proto-oncogene. | |
| GN | XMRK. | |
| OS | Xiphophorus xiphidium. | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC | Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; | |
| OC | Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; | |
| OC | Cyprinodontiformes; Poeciliidae; Xiphophorus. | |
| OX | NCBI_TaxID=8086; | |
| RN | [1] | |
| RP | SEQUENCE FROM N.A. | |
| RC | STRAIN=RIO PURIFICATION; | |
| RX | MEDLINE=98241172; PubMed=9582016; | |
| RA | Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J., | |
| RA | Altschmied J., Schartl M.; | |
| RT | "Activation of the Xmrk proto-oncogene of Xiphophorus by | |
| RT | overexpression and mutational alterations."; | |
| RL | Oncogene 16:1681-1690(1998). | |
| RN | [2] | |
| RP | SEQUENCE FROM N.A. | |
| RC | STRAIN=RIO PURIFICATION; | |
| RA | Schartl M.; | |
| RL | Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases. | |
| DR | EMBL; U53471; AAD10500.2; .. | |
| DR | HSSP; P11362; 1FGK. | |
| DR | InterPro; IPR000345; CytC_heme_bind. | |
| DR | InterPro; IPR000494; EGFR_L_domain. | |
| DR | InterPro; IPR000719; Euk_pkinase. | |
| DR | InterPro; IPR002174; Furin-like. | |
| DR | InterPro; IPR001899; Gram_pos_anchor. | |
| DR | InterPro; IPR001245; Tyr_pkinase. | |
| DR | Pfam; PF00757; Furin-like; 1. | |
| DR | Pfam; PF00069; pkinase; 1. | |
| DR | Pfam; PF01030; Recep_L_domain; 2. | |
| DR | PRINTS; PR001109; TYRKINASE. | |
| DR | ProDom; PD000001; Euk_pkinase; 1. | |
| DR | SMART; SM00261; FU; 3. | |
| DR | SMART; SM00219; TyrcK; 1. | |
| DR | PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1. | |
| DR | PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN 1. | |

| | |
|--|--|
| Best Local Similarity 46.9%; Pred. No. 1.6e-195; Matches 538; Conservative 172; Mismatches 351; Indels 85; Gaps 26; | |
| QY 161 LCYQDTILWKDIFKKNQALALTLIDTNRSRACHPCSPMKSGRSGESSEDQSLTRTV 220 | |
| DB 3 LCFADTHHQDVRNPASNFVLPTNGSSGGRCHKCTG--RCWGPTENHCQTILTKVC 61 | |
| QY 221 AGGC--ARCKGRLPTDCHEQAAGCTGPKHSCLACLFPHNSGICELHCP--ALVFNFTV 278 | |
| DB 62 AEQDRCYGVVSDCHREAGCGGPKDTCFACMFNDSGACVTCQPTFVNPPTF 121 | |
| QY 279 SFMLRVPKVSASHLERYTFGACVTCACPVNVLSTDVGSCTLVCLPHNQEVTAEDGTQCE 338 | |
| DB 122 -----OLEHNNAKYTYGACVKKCPHNFV--VDSNCVACRACPSKMEV--BENGIMCK 172 | |
| QY 339 KSKPCARVCYGLGMEHLREVPAVTSANIQEFAGCKIFGSLAFLPESPDCGPASNTAPL 398 | |
| DB 173 PCTDICPKACDGGTGSLSVAQTVDSSNIDKFNCTKINGNLFLVGTGHGDPYHTIAAI 232 | |
| QY 399 OPEQLQVETLEEITGYLISAWPDLPSLVSFQNLQVIRILHNGAYSLTLQGLGISW 458 | |
| DB 233 NPEKLNIFQTVREITGYLNIQSWPENMTDFRVFSNLVTIGGRALYSGLSLLILKQGGITS 292 | |
| QY 459 LGLRSLRLGSLALIHNNTHLCFVHTVPWDLFRNPHQALHTANRPEDECVGEGGLACH 518 | |
| DB 293 LQFQSLKQISAGNIVITDNLNLCYHVTNMTSLFSTPSOKTVIHRNKAENCTADGMVCN 352 | |
| QY 519 OLCARGHGWGPTOCVNCQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECP--QN 577 | |
| DB 353 ELCSDDGCGPDDCLSKRIRGTCIESNLYDGEFREFANGSVCMEDPQCEKMBD 412 | |
| QY 578 GSVTCFGEADQCAHAKDPFCFVARCPGVKPDLSYMPWKFPDDEGACQCPINCT 637 | |
| DB 413 NMITCYGPGDHTCKCFHFQDGNVCEKCPDGLQGANSE--IFKYADEDECHPCPNCT 470 | |
| QY 638 HSCVDLDKGC-----PAORASPLTSIVSAVV--GILLVVLGVVFGILIKRRQ 685 | |
| DB 471 QGCRGPASHDCIYYPTWRTSLPQAR--TPL--IAAGVIGGLFIIIMGLTEAVVRRKS 527 | |
| QY 686 OKIRYWRRLLOETELVEPLTPSGAMPNQAMRILKETLRKVKLGSAGTGVYKGM 745 | |
| DB 528 IK--KRAURRFL--ETELVEPLTPSGTAPNQALRIKETLRKVVLGSGAGTGVYKGM 585 | |
| QY 746 IPDGENVIPVAIKVLENTSPKANKEILDEYVMAVGSPVSVRLGLICLTSTVOLVTO 805 | |
| DB 586 VPEGETVIPVAIKILNETTGPKANVEFMDEALIMASMDHPHLVRLGLVCLFTQLVTO 645 | |
| QY 806 LMPYGLLDHVRNRLGSLQDLNLCMKQIAKMSYLEDVRLVLRDLAARNVLKSPNHV 865 | |
| DB 646 LMPHGCLLDYVHEHKDNLGSLNLCVQIAKGMYLEERLVRDLAARNVLKSPNHV 705 | |
| QY 866 KITDFGLARLLDIDETVHADGKVPKIMMALESILRRFTHOSDVWSYGVTVWELMTFG 925 | |
| DB 706 KITDFGLARLLDDEKEYNADGKVPKIMMALECIHYRKFTHOSDVWSYGVTVWELMTFG 765 | |
| QY 926 AKPYDGIIPAREIPDLLEKGERLPQPPICITIDYVMIMVKCMWIDSECRPRFELVSEFSRM 985 | |
| DB 766 GKPYDGIIPREIPDLLEKGERLPQPPICITIDYVMIMVKCMWIDADSRPKELAAEFSEFM 825 | |
| QY 986 ARDPORFVVIQNE--LGPAFLPDSTFYRSLLEDMDGDLVDAEYLVPOQGFCDPDPAPG 1044 | |
| DB 826 ARDPORYLVIQDDRMLKPLSPNDSKFFQNLLEDLEDMDMAEYLVLP--QAFNIPPIYT 884 | |
| QY 1045 AGGMVHRRHSSTSGGDLTLGLEPSEEEAPRS--PLAP--SEGAGSDVFDGDLGMGAA 1101 | |
| DB 885 SSTRIDSNRNQVYRDGGVAEQGV--PMYRAPGCIIPAPVAQGATAEIFEDTCCNGTL 943 | |
| QY 1102 KGLOSLPHDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQEPVYNQPDVRPQ 1154 | |
| DB 944 RKQVATLAKEDSSSTQRYSDPTVFIPERVIRCELDEDEGYMTPMRKPTDYLNPVEENPF 1003 | |
| QY 1155 PPSREGPLPAA--RPAGATLERAKTLSPCKNGVVKDVF-----AFGGAIVENPEYLTPT 1205 | |
| | |
| DB 1004 VSRRKNGDLQAVDNPETHN-----APNGOPKAEDYVNEPLYLNTFANTLENAEYL-- 1054 | |
| QY 1206 OCGAARQPHPPAFSPADNLVYWDODPERCA--PPSTFKGTPT-----A 1249 | |
| DB 1055 -----KNLPEKAKAFDNPDMYHNSUPPRSTLQHPDYLOEYSTKYFYKQNGRIRPIVA 1108 | |
| QY 1250 ENPEYL 1255 | |
| DB 1109 ENPEYL 1114 | |
| RESULT 6 | |
| ID P79754 PRELIMINARY; PRT: 1328 AA. | |
| AC P79754; | |
| DT 01-MAY-1997 (TREMBLrel. 03, Created) | |
| DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update) | |
| DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update) | |
| DE ErbB3. | |
| OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes). | |
| OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; | |
| OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; | |
| OC Tetraodontidae; Takifugu. | |
| OX NCBI_TaxID=31033; | |
| RN [1] | |
| RP SEQUENCE FROM N.A. | |
| RX MEDLINE=99177347; PubMed=10077531; | |
| RA Gellner K., Brenner S.; | |
| RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu | |
| RT rubripes."; | |
| RL Genome Res. 9:251-258(1999). | |
| DR EMBL; AF056116; AAC34391.1; -- | |
| DR HSSP; P11362; IFGK. | |
| DR InterPro; IPR000494; EGFR_L domain. | |
| DR InterPro; IPR000719; Euk_pkinase. | |
| DR InterPro; IPR002174; Furin-like. | |
| DR InterPro; IPR001245; Tyr_pkinase. | |
| DR Pfam; PF00757; Furin-like; 1. | |
| DR Pfam; PF00069; pkinase; 1. | |
| DR Pfam; PF01030; Recep_L domain; 2. | |
| DR ProDom; PD000001; Euk_pkinase; 1. | |
| DR SMART; SM00261; FU; 3. | |
| DR SMART; SM00219; TyrcK; 1. | |
| DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1. | |
| DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1. | |
| KW ATP-binding; Transferase. | |
| SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64; | |
| Query Match | |
| Best Local Similarity 33.4%; Score 2284; DB 13; Length 1328; | |
| Matches 521; Conservative 157; Mismatches 417; Indels 198; Gaps 33; | |
| QY 9 WGLLLALAPP--GAASTQ----VCTGDMKRLRASPETHLDMRLHLYGQCQVVOGNLEL 62 | |
| DB 4 WRLILMCVASRLRAASSQTQEAFCPTGNLSSQTQENQYNLNKDRYKGCIEIMGNLEI 63 | |
| QY 63 TVLPTNASLSFLDITQEVGYVLIHQNQVRVPLORLIRVGTQLPFDNYALAVLNDGDP 122 | |
| DB 64 TQIESNWDPSFLKTRIVTYVYLIAMNHFEIPLQGLRVIRGNSLYERRFALLSVFLN--- 120 | |
| QY 123 LNNTPVTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHKNNQLALT 182 | |
| DB 121 ----YPKDG--PSGNLQGLMNLTEILDGGVLIINNKYLRYPGWVYWRDII--RNNDAPIE 173 | |
| QY 183 LIDTNRSRACHPCSPMKSGRSGESSEDQSLTRTVCAAGC--ARCKGRLPTDCCHCEQA 241 | |
| DB 174 IQFNGERGVCV---KSC--GNYCWGFGKQCOQLTKTVCAPOCNDRCFGTSPRDCCHIECA 229 | |
| QY 242 AGCTGPKHSCLACLFPHNSGICELHCP--ALVFNFTVSVFMRVPKVSASHLERYTFGAS 300 | |
| DB 230 AGCKGPLDTCFACRLFNDSGACVPCQPTLIYNKQTF-----QMETNPNAKYQYGS 282 | |


```
QY 430 VFQNLQVIRGILHNGAY-SLTQGLGISWLGRLSRLRELGLALIHNNTHLCFVHTVPW 488
Db 384 YERNLEVGGRLKENLFPASVIVKTSLSKLSKRVNSGIVILENSDLCFVEDIDM 443
QY 489 DOLFNPQALHTANRBEDECVGEGLAGHQICARGHCWGPQTQVNCVCSQFLRQOECVE 548
Db 444 SEIKSSDHEVNVQNRNATEHEGMECEQSCAGKMGKGPQEQCLECKNVKYGKCLID 503
QY 549 ECRVLQGLPREY-VNARHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFPCVAPCP 607
Db 504 SKC---SLPRLYSVDSKTCGDCHQECID-----FCYGPNEDCGSMNVKQGRFCVAECP 555
QY 608 -----SGLVPLQEEGPKQLSGKAVCRKCHPRCKKCTGYGFHEQFCQECTGYKKGEQCECPQD 624
Db 556 TTKHAMNGTGINCHKTCVCGCRPRDTIAPDGCISCDAKIIIGSDAKIERCLMKDESCPDGY 615
QY 625 -----REG----- 627
Db 616 YSDYVLQEEGPKQLSGKAVCRKCHPRCKKCTGYGFHEQFCQECTGYKKGEQCECPQD 675
QY 628 -----ACQPCPINCT-----HSCVDL-----DD-----KGCPEAQ--- 652
Db 676 FYANETRCLPCHQECRCCHGLGDHHECRNLKLFEGDPYDNATFTCVSNCPASHPYK 735
QY 653 -----RASPLTSIVSAVVGILLVVLGVVFGI---LIKRRQOK 687
Db 736 RFPQEAQKIGPYCSADSMQSLRIEPTQVKIVMSGVMALILLVFGIAFLVFSRHKNK 795
QY 688 IRKYTHRRLLQETELVEPTPSGAMPNQAMRILKETELRKVKVLGSGAFGTVYGIWIP 747
Db 796 KDAVKMTALACEDSEPLRPSNVGNLTKLRIKEAEIRRGVGLMGAGFRVFKGVMP 855
QY 748 DGENVKIPVAIKVLRENTSPKANKEITLDEAYVMAGVSPYVSRLLGICLTSTVQVLTQLM 807
Db 856 EGESVKIPVAIKVLEMSGESKEFLEAYIMASVEHPNLKLLKLAACVMTSQMMLITQLM 915
QY 808 PYGCLLDHVNRGRGSDLLNWCMIQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKI 867
Db 916 PLGCLLDVVRNKKIGSKALLNWSQIARGMAYLEERLVRDLAARNVLVQTPSCVKI 975
QY 868 TDFGLARLLDIDETEHADGGKVPKWMALSIILRRFTHQSDVWSYGVTVWELMTFGAK 927
Db 976 TVFGLAKLLDFDSDEYRAAGGKMPIKWLALCIRHRVFTSKDVMAFGITIWELLTYGAR 1035
QY 928 PYDGPAREIPDLLEKGERLPPOPICTIDVYIMVKKWIMIDSECRPRPRELVSEFSMAR 987
Db 1036 PYENVPKDVPELIEIGHKLPOPDICSLDVYICILSCWVLDADARTFKQLAETFAEKAR 1095
QY 988 DPQRFVIONEDLGASPLDSTFYRSLLEDDEDDMDGLV----- 1024
Db 1096 DPGRYLMI-----PGDKFMRPLPSYTNQDEKDLIRTLAPVMAAAAAAAGASNVYD 1146
QY 1025 -----DABEYLVPQGFPCPDPAFCAGGMVHRRHSSTRSGGGDLTLGLFSEEEAPR 1078
Db 1147 VPSITAEITDLOPKTRPSIMLPGPSA-----VEPS-DEMPK 1182
QY 1079 S-----PLAP---SEGAGSDVFDGLGMAAKQLSLPDPSPLORYSEDPTVPLPSE 1129
Db 1183 SLRYCKDPLKPDDETDGHEV-----GVGGIR-----LNLPID 1216
QY 1130 TDGYVAPLTCSPQPEYVNPQDVRPPSPREGFLPAARPAGATLERAKTLSPGKNGVWKD 1189
Db 1217 EDDYLMPTCQSQ---NQS-----TFG---YND 1237
QY 1190 VFAFGAVENPEYL-----TPQGGNAOPHPHPAPFADNLYYWDQPPBERGAPP 1240
Db 1238 LIGVPSVDNPEYLMGMSQAIAGLAQSGMG---PHTPP-----PP 1274
QY 1241 STFKGTPTAENPE 1253
Db 1275 NTPNGMPTHQHSQ 1287
```

RESULT 8

Q9UK79

ID Q9UK79 PRELIMINARY; PRT; 419 AA.

AC Q9UK79;

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Herstatin.

GN HER-2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9415951; PubMed=10485918;

RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;

RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted

RT autoinhibitor";

RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).

RN [2]

RP SEQUENCE FROM N.A.

RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF177761; AAD56009.2; ..

DR InterPro; IPR000494; EGFR_L domain.

DR InterPro; IPR002174; Furin-Like.

DR Pfam; PF00757; Furin-Like; 1.

DR Pfam; PF01030; Recep_L domain; 1.

DR SMART; SM00261; FU; 1.

SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347B2D030C CRC64;

Query Match

Best Local Similarity 25.9%; Score 1770.5; DB 4; Length 419;

Matches 340; Conservative 7; Mismatches 40; Indels 13; Gaps 3;

```
QY 1 MELALCRWGLLLALLPGGAASQVCTGTDMLRLPASPEHLMLRHLVYQGVVQGNL 60
Db 1 MELALCRWGLLLALLPGGAASQVCTGTDMLRLPASPEHLMLRHLVYQGVVQGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQGYVLIQAHNQVQPLQRLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIOEVQGYVLIQAHNQVQPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGVLTQRPOLCYQDTILWKDIFHKQNLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGVLTQRPOLCYQDTILWKDIFHKQNLA 180
QY 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACARCKGLPTDCCHEOC 240
Db 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACARCKGLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVFNFTVSWFLRVPKVSASHLERYTFGAS 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVFNFTVSWFLRVPKVSASHLERYTFGAS 300
QY 301 CVTACPNYLSLDVSCSTLVCPHLHNOEVTAEQGTQCEKCKPCARVCYGLGHEHLREVR 360
Db 301 CVTACPNYLSLDVSCSTLVCPHLHNOEVTAEQGTQCEKCKPCARVCYGLGHEHLREVR 360
QY 361 AVTSANIQEFAGCKKIFGSLAFPLPSFDGDPASNTAPLOP 400
Db 350 AAVPVLKMQPG--PAHPVLSFLRPSWDLVSFAFSLPLAP 387
```

RESULT 9

Q8R2X1

ID Q8R2X1 PRELIMINARY; PRT; 367 AA.

AC Q8R2X1;

DT 01-JUN-2002 (Tremblrel. 21, Created)

DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Hypothetical 40.2 kDa protein.

OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC027080; AAH27080.1; -;
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 3.3e-123; Indels 0; Gaps 0;
Matches 323; Conservative 11; Mismatches 33;

QY 895 MALESILRRRTHQSDVMSYGVTVVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPIC 954
Db |||||
QY 955 IDVYIMVWCMWIDSECRPRPRELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSL 1014
Db |||||
QY 1015 LEDDDMGDLVDAEEYLVPOQGFCDPAPAGAGMVHRRSSSTRSGGDLTLGLEPSEE 1074
Db |||||
QY 1075 EAPRSLAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVLPSPSETDGV 1134
Db |||||
QY 1135 APLTCSPOEYVNPQDVPQPPSPREGPLPAARAGATLERAKTILSPGKNGVVKDVFAG 1194
Db |||||
QY 1195 GAVENPEYLTPOGGAPOPHPPAPFSPAFDNLVYWDQPPERAGPSTFKGTPTAENPEY 1254
Db |||||
QY 1255 LGLDVPV 1261
Db |||||
QY 361 LGLDVPV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
evolution of distinct viral genomes carrying mutant v-erbB genes with
different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -;
DR HSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.

DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TVR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFEID63 CRC64;

Query Match 25.1%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 2.5e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 575 PONGSVTCFGEADOCVACAHYKDPFCVAPCPSPGVKPDLSYMPIWKFDPDEEGACPCPI 634
Db |||||
QY 635 NCTHSCVDLDDDDCAEPQASPLTSIVSAVV-GILLVVVLGVVFGVGLIKRRQOKIRKYM 693
Db |||||
QY 198 NCTRCCKPGLEGCP---NGSKTPSTAAGVVGGLLCLVVGGLGIGLYLRRR-HIVKRTL 253
QY 694 RLLQETELVEPLTPSGAMPNQAOMRIKTELKRVKVLGSGAFGVYKGIWIPDGENVK 753
Db |||||
QY 254 RLLQERELVEPLTPSGEAPNQAHLRIKTEFKVKVVLGSGAFGVYKGLWIPGEKVK 313
QY 754 IPVAIKVLRNTSPKANKEILDEAYVMAGVSPYSRLIGICLTSTVQLVTQLMPYGCLL 813
Db |||||
QY 314 IPVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLIGICLTSTVQLVTQLMPYGCLL 373
QY 814 DHVRENRLGSLGODLLNMCWIAKMSVLELDVLRDLAARNLVKSPHVKITDFGLA 873
Db |||||
QY 374 DYIREHKONIGSQYLLNMCVQIAKGMVLEERLVRDLAARNLVKTPQHVKITDFGLA 433
QY 874 RLLDIDETEHADGGKVPKMALESILRRRTHQSDVMSYGVTVVWELMTFGAKPYDGP 933
Db |||||
QY 434 KLGADEYHAEAGKVPKMALESILHRIYTHQSDVMSYGVTVVWELMTFGSKPYDGP 493
QY 934 AREIPDLLEKGERLPQPPICITDVTVMWCMWIDSECRPRPRELVSEFSRMARDPQRFV 993
Db |||||
QY 494 ASEISSVLEKGERLPQPPICITDVTVMWCMWIDADSPPKPRELIAEFSKWARDPPYL 553
QY 994 VIQ-NEDLGPASPLDSTFVRSILLEDDMGDLVDAEEYLVPOQGFCDPAPAGAGMVHHR 1052
Db |||||
QY 554 VIQDERVHLPSPDTSKFYRTLMEBEDMEDIVDADEYLVPHQGF-----598
QY 1053 HRSSTRSGGDLTLGLEPSEEEAPRSL-----APSEGAGSDVFDGDLGMAAGLQSL 1107
Db |||||
QY 599 -NSPST-----SRPLSLSLATSNSATNCID-----RNGQGH 631
QY 1108 PTHDPSLQRYSEDPTVLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLA 1165
Db |||||
QY 632 PVREDSFVQRYSSDPTGNFLEESIDGFL-----PAPEYVQ--LMPKKPS-----675
QY 1166 ARPAGATLERAKTILSPGKNGVVKDVF-----AFGAVENPEYL 1203
Db |||||
QY 676 -----TAMVQNIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 715

RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,


```
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; IFKGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MM; C6D9CBA7ADF725E1 CRC64;

Query Match 25.1%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 2.5e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 584 GPEADQCVACAHYKDPFCVACPSGVKPDLSYMPIWKFPDBEGACQPCPINCTHSCVDL 643
DB 1 GP-DHCKCAHFIDGPHCVKACPAGVGENDTL-VKYNADANAVCOLCHPNCTRCCKGP 57

QY 644 DKGCPAQRASPLTISVSAV-VGILLVVLGVVFGILIKRQOKIRKYMRLLOETEL 702
DB 58 GLEGCP-NGSKTPSIAAGVGGGLVWGLGIGLYLRR-HIVRKRTLRLLOREL 113

QY 703 VEPLTPSGAMPQAOIRLKETELRKVKVLGSGAGFTVYKGIWIPGENVKIPVALKVR 762
DB 114 VEPLTPSGAPQAHURLIKETEFKKVKVLGSGAGFTVYKGLWIPGEKVKIPVALKEUR 173

QY 763 ENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVTLQMPYGCLLDHRVNRGR 822
DB 174 EATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVOLITQMPYGCLLDVIRESKON 233

QY 823 LGSODLLNWCMOIAKMSVLEDRVLVHRLDAAARNVLKSPNHVKITDFGLARLLDIDETE 882
DB 234 IGSQYLLNWCVOIAKGMNLYEBRRLLVHRLDAAARNVLVKTPOHVKITDFGLAKLLGADEKE 293

QY 883 YHADGKVPKWMALLESILRRRFTHOSDVMSYGVTVWELMTFGAKYDGI PAREIPDLLE 942
DB 294 YHAEGKVPKWMALLESILHRYTHOSDVMSYGVTVWELMTFGSKYDGI PASEISSVLE 353

QY 943 KGERLPQPICTIDVYMIWVKCMIIDSECRPRFRELVSFESRMARDPQRFVVIQ-NEDLG 1001
DB 354 KGERLPQPPICTIDVYMIWVKCMIIDADSRPKFRELIAEFSKWARDPPRYLVVIQDERMH 413

QY 1002 PASPLDSTYRSILLEDDMGDLVDAEYLVPOQGFCDPAPGAGGMVHRRSSSTRSG 1061
DB 414 LPSPTDSKYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454

QY 1062 GGDLTGLGPESEEA PRSPL-----APSECAQSDVFDGDLGMAAKGLQSLPHDPSPIQ 1116
DB 455 -----SRTLLSLSATSNNNSATNCID-----RNGQGHVPREDSFVQ 491

QY 1117 RYSEPTVPLPSET--DGVVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARAGATLE 1174
DB 492 RYSSDPTGNFLEESIDDGFL-----PAPEYVNQ--LMPKKS----- 526

QY 1175 RAKTLPNGKGVVVKDVF-----AFGAVENPEYL 1203
DB 527 ----TAMVQNYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL 566

RESULT 12
Q8WYV0
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;

QY 895 MALESILRRRFTHOSDVMSYGVTVWELMTFGAKYDGI PAREIPDLLEKGERLPQPICT 954
DB 1 MALESILRRRFTHOSDVMSYGVTVWELMTFGAKYDGI PAREIPDLLEKGERLPQPICT 60

QY 955 IDVYMIWVKCMIIDSECRPRFRELVSFESRMARDPQRFVVIQ-NEDLGASPLDSTFYRSL 1014
DB 61 IDVYMIWVKCMIIDSECRPRFRELVSFESRMARDPQRFVVIQ-NEDLGASPLDSTFYRSL 120

QY 1015 LEDDDMGDLVDAEYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLPSE 1074
DB 121 LEDDDMGDLVDAEYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLPSE 180

QY 1075 EAPRSLAPSEGAGSDVFDGDLGMAAKGLQSLPHDPSPLORYSBDPTVPLPSETDGYV 1134
DB 181 EAPRSLAPSEGAGSDVFDGDLGMAAKGLQSLPHDPSPLORYSBDPTVPLPSETDGYV 240

QY 1135 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTLPNGKGVVVKDVPFG 1194
DB 241 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAARAGATLERPKTLPNGKGVVVKDVPFG 300

QY 1195 GAVENPEYLTPOGGAAPQ-----HPPA---FSPAFDNL 1226
DB 301 GAVENPEYLTPOGGAALSFTLLPSAQPSSTTITGTRTHQSGGLHPAPSKGHLRQRTQST 360

QY 1227 YWVD-QDPPER-----GAPSTFKGTPTAEN 1251
DB 361 WWTQCPEPGQVRRSPDVSSGREGLTSAGIKRWEGPPTTSRGTCARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN GAG, v-ERB-A, v-ERB-B.
OS Avian erythroblastosis virus.
```

Vituses; Retrovirdae; Retrovirdae; Avian type C retrovirdae.

```
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruskon A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RL transforming potential of the oncogene v-erb-B.";
RL Oncogene 5:15-24(1990).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Steroid_receptor.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHOMER.
DR PRINTS; PR00047; STROIDFINGER.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR PRODOM; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00219; TyrKc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791EAE95CE CRC64;

Query Match 24.2%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 4.1e-116;
Matches 358; Conservative 73; Mismatches 149; Indels 119; Gaps 18;

QY 547 VECCRVQLGLPRE-VYNAR-HCLP-----CHPEQC 574
DB 354 IEKQESYLLAFAYIYINRGKINIPHWSKLLMKVADLRMIGAYHASRFLHMKVECPTELS 413
QY 575 PONGSVTCFGEADQCVACAHYKDPFCVACRCPGKPDLSYMPWKFPDEEGACQCP1 634
DB 414 PQE-----VGP--DHCNKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHP 465
QY 635 NCTHSCVDLDDKGPACORASPLTSIVSAVY-GILLVVVLGVVFGILIKRQKIRKYTM 693
DB 466 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVGIGLGLYLR--HIVRKRTL 521
QY 694 RLLOETLEVEPLTPSGAMPNOAQRILKETELKVKVKGSGAGTVYKGIWIDGENVK 753
DB 522 RLLQERELVEPLTPSGEAPNOAHLRLKETEFKKVKVKGFGAGTVYKGIWIDPEGEKVT 581
QY 754 IPVAIKVLRNTPSKANKEILDEAYVMAGVSGPVVSRLLGICLTSTVQLTQMPYGCGLL 813
DB 582 IPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLTQMPYGCGLL 641
QY 814 DHVRENRLGSGDILLNWCQIAKGMYSLEDRVLVHRDLAARNVVKSPNHVKITDFGLA 873
DB 642 DYIREHKNIGSQYLLNWCQIAKGMNLEERHVRDLAARNVVKTPQHVKITDFGLA 701
QY 874 RLIDIDETEYHAGDGKVPKIMWALESILRRFTHQSDVWSVGVTVWELMTFGAKPYDGIP 933
DB 702 KQLGADKEKHAEGGKVPKIMWALESILHRYTHQSDVWSVGVTVWELMTFGSKPYDGIP 761
QY 934 AREIPDLLEKGERLPQPPICITIDVYIMVWKCMYIDSECRPRFRELVSFESRMARDPORFV 993
DB 762 ASEISSVLEKGERLPQPPICITIDVYIMVWKCMWNSGADSRPKFRELIAEFSKWARDPPRYL 821
```

```
QY 994 VIQ-NEDLGASPLDSTFYRSLLDEDDMGDLVDABEYLVPQOQFFCFDPPAGAGVMVHR 1052
DB 822 VIQDERMHLPSPTDSKFYRTLMEBEDMEDIVDAEYLVPHQGF-----866
QY 1053 HRSSSTRSGGDLTLGLPESEEEAPRSLAPSEGAGSDVFDGDLGMAAKGLOSLEPHDP 1112
DB 867 -NSPST-----SRTPLLSLSATS-----NSATKCIDRNGGH--898
QY 1113 SPLQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYVNOQDVVRPQPPSPREGPLPAARAGAT 1172
DB 899 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQONQIYINISLT 936
QY 1173 -LERAKTLSPGKGVVVDVFAFGGAVENPEYL 1203
DB 937 AISKLPMDSRVQN-----SHSTAVDNPEYL 961

RESULT 14
Q85468
ID Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retrovirdae; Retrovirdae; Avian type C retrovirdae.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RL virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CAA30024.1; -
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CCA0F8AF4 CRC64;

Query Match 24.1%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 7.8e-116;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 584 GPEADQCVACAHYKDPFCVACRCPGKPDLSYMPWKFPDEEGACQCPINCTHSCVDL 643
DB 1 GP--DHCNKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57
QY 644 DDKCPAEQASPLTSIVSAVY-GILLVVVLGVVFGILIKRQKIRKYTHRLLQETEL 702
DB 58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVGIGLGLYLR--HIVRKRTLRLQLEREL 113
QY 703 VEPLTPSGAMPNOAQRILKETELKVKVKGSGAGTVYKGIWIPDGENVKIPVAIKVL 762
DB 114 VEPLTPSGEAPNOAHLRLKETEFKKVKVKGFGAGTVYKGIWIPDGENVKIPVAIKEL 173
QY 763 ENTSPKANKEILDEAYVMAGVSGPVVSRLLGICLTSTVQLTQMPYGCGLLDYIREHKN 822
DB 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLTQMPYGCGLLDYIREHKN 233
QY 823 LGSODILLNWCQIAKGMYSLEDRVLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETE 882
DB 234 IGSQYLLNWCQIAKGMNLEERHVRDLAARNVVKTPQHVKITDFGLAKQLGADKE 293
```


Search completed: July 22, 2003, 09:00:55
Job time : 53.5887 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-250-264-14

Perfected score: 6939

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq 101002.*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 6708 | 98.1 | 1255 | 21 | Human heregulin 2 |
| 2 | 6708 | 98.1 | 1255 | 22 | Human tyrosine kin |
| 3 | 6708 | 98.1 | 1255 | 22 | HER2 transgene pla |
| 4 | 6708 | 98.1 | 1255 | 23 | Human HER2 (ErBB2) |
| 5 | 6702 | 98.0 | 1255 | 17 | HER-2/neu protein. |
| 6 | 6702 | 98.0 | 1255 | 20 | Human HER-2/neu on |
| 7 | 6702 | 98.0 | 1255 | 21 | Human HER-2/neu pr |
| 8 | 6702 | 98.0 | 1255 | 21 | Amino acid sequenc |
| 9 | 6702 | 98.0 | 1255 | 22 | Human HER-2/neu pr |
| 10 | 6702 | 98.0 | 1255 | 22 | HER2/neu amino aci |

| | | | | | | |
|----|--------|------|------|----|----------|--------------------|
| 11 | 6702 | 98.0 | 1255 | 23 | AAE24067 | Human Her-2 protei |
| 12 | 6702 | 98.0 | 1255 | 23 | AAE20479 | Human Her-2/neu pr |
| 13 | 6702 | 98.0 | 1255 | 23 | AAE20479 | Human Her-2/neu pr |
| 14 | 6702 | 98.0 | 1255 | 23 | AAU77114 | Human Her-2/neu po |
| 15 | 6659 | 97.4 | 1433 | 14 | AAE39568 | Sequence of c-erbB |
| 16 | 6538 | 95.6 | 1223 | 23 | AAU98923 | Human breast cance |
| 17 | 6385 | 93.4 | 1200 | 21 | AAE21208 | Human HER-2/neu pr |
| 18 | 5905.5 | 86.4 | 1256 | 21 | AAE21199 | Rat Her-2/neu pr |
| 19 | 5905.5 | 86.4 | 1256 | 21 | AAE21199 | Rat Her-2/neu pr |
| 20 | 5881.5 | 86.0 | 1256 | 21 | AAE21206 | Mouse Her-2/neu pr |
| 21 | 5881.5 | 86.0 | 1256 | 21 | AAE21206 | Mouse Her-2/neu pr |
| 22 | 5881.5 | 86.0 | 1256 | 21 | AAE21206 | Amino acid sequenc |
| 23 | 4788 | 70.0 | 919 | 21 | AAE21203 | Mouse Her-2/neu on |
| 24 | 4788 | 70.0 | 919 | 21 | AAE21203 | Human HER-2/neu fu |
| 25 | 4031.5 | 58.9 | 920 | 23 | AAE21148 | Her-2/neu extracel |
| 26 | 4031.5 | 58.9 | 926 | 23 | AAE21152 | Mouse Her-2/neu ex |
| 27 | 3672 | 53.7 | 712 | 21 | AAE21204 | Human HER-2/neu ex |
| 28 | 3672 | 53.7 | 712 | 21 | AAE21204 | Human HER-2/neu fu |
| 29 | 3526 | 51.6 | 782 | 18 | AAE19764 | Her-2/neu extracel |
| 30 | 3524 | 51.5 | 653 | 21 | AAE21200 | Her2-QM-CSF immuno |
| 31 | 3524 | 51.5 | 653 | 21 | AAE21200 | Extracellular HER- |
| 32 | 3486 | 51.0 | 645 | 22 | AAE20481 | Human ErbB2 oncopr |
| 33 | 3486 | 51.0 | 645 | 22 | AAE20481 | Human ErbB2 extrac |
| 34 | 3421 | 50.0 | 951 | 21 | AAE44993 | DC8scFv-erbB2EC fu |
| 35 | 3318 | 48.5 | 624 | 11 | AAE08222 | Amino acid sequenc |
| 36 | 3116 | 45.6 | 1210 | 21 | AAE19259 | Human EGF receptor |
| 37 | 3116 | 45.6 | 1210 | 21 | AAE19259 | Human EGF receptor |
| 38 | 3116 | 45.6 | 1210 | 23 | AAE23019 | Human Her-1 protei |
| 39 | 3116 | 45.6 | 1210 | 23 | AAE23019 | Human epidermal gr |
| 40 | 3114 | 45.5 | 1210 | 22 | AAE20483 | Amino acid sequenc |
| 41 | 3084 | 45.1 | 583 | 23 | AAE20483 | Human protein for |
| 42 | 3084 | 45.1 | 583 | 23 | AAE20481 | Human protein for |
| 43 | 3083 | 45.1 | 589 | 23 | AAE20484 | Human protein for |
| 44 | 3083 | 45.1 | 600 | 23 | AAE20482 | Human protein for |
| 45 | 3075 | 45.0 | 1210 | 23 | ABP51768 | Human epidermal gr |

ALIGNMENTS

RESULT 1
AAE24067
ID AAY92620 standard; Protein; 1255 AA.
XX .
AC AAY92620;
XX .
DT 10-AUG-2000 (first entry)
XX .
DE Human heregulin 2 (Her2).
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX .
OS Homo sapiens.
XX .
FH Key Location/Qualifiers
FT Domain 1..173
FT /label= N-terminal
FT /note= "mature polypeptide"
FT Region 5..25
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 59..73
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 103..117
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 149..163
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Domain 174..323

FT Region /label= Cysteine_rich_domain
 210..224
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 250..264
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 324..483
 /label= Ligand_binding_domain
 325..339
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 369..383
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 465..479
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 484..623
 /label= Cysteine_rich_domain
 579..593
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 624..654
 /label= Transmembrane_domain
 632..652
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 653..667
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 655..1010
 /label= Tyrosine_kinase_domain
 661..675
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 695..709
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 710..730
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 1011..1235
 /label= C-terminal_domain
 WO200020027-A2.
 13-APR-2000.
 05-OCT-1999; 99WO-DK00525.
 05-OCT-1998; 98DK-0001261.
 20-OCT-1998; 98US-0105011.
 (MEBI-) M & E BIOTECH AS.
 Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 Gautam A, Birk P, Karlsson G;
 WPI; 2000-349917/30.
 N-PSDB; AAA09455.
 Inducing immune responses to weakly immunogenic, tumor associated
 peptide antigens for the treatment of breast and prostate cancer
 Claim 62; Page 193-198; 220pp; English.
 This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 Her2 can be used in the claimed method as an autovaccine to induce a CTL
 response. Subdominant CTL epitopes, antibody binding regions and
 cysteine residues involved in disulfide bonds are preserved in the
 immunogenized forms. Regions suitable for the insertion of foreign T
 helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX
 SQ Sequence 1255 AA;
 Query Match 98.1%; Score 6708; DB 21; Length 1255;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 1242; Conservative 2; Mismatches 11; Indels 6; Gaps 1;
 QY 1 MELAALCRWGLLLALLPFGAASQTCTGTDMLRLPASPTHLDMLRHLQYQCVVQGNL 60
 Db
 1 MELAALCRWGLLLALLPFGAASQTCTGTDMLRLPASPTHLDMLRHLQYQCVVQGNL 60
 QY 61 ELTYLPTNASLSFLQDIOEVQYVLI AHNOYRQVPLQRLRIVRGTLQEDNVALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIOEVQYVLI AHNOYRQVPLQRLRIVRGTLQEDNVALAVLDNG 120
 QY 121 DPLNNTTPTVGASPGCLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
 Db 121 DPLNNTTPTVGASPGCLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
 QY 181 LTLIDTNRSRACHPCSKGSRMGESSEDCQSLTRTVACGACARCKGPLPTDCHEQC 240
 Db 181 LTLIDTNRSRACHPCSKGSRMGESSEDCQSLTRTVACGACARCKGPLPTDCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVFNFTVSFWLRVPKVSASHLERTFFGAS 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVFNFTVSFWLRVPKVSASHLERTFFGAS 300
 QY 301 CVTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTORCEKSCPKARVCYGLGMEHLREVR 360
 Db 295 CVTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTORCEKSCPKARVCYGLGMEHLREVR 354
 QY 361 AVTSANIOEFAGCKKIFGSLAPLPSFDGDPASNTAPLOPEQLQVFETLEETGYLYISA 420
 Db 355 AVTSANIOEFAGCKKIFGSLAPLPSFDGDPASNTAPLOPEQLQVFETLEETGYLYISA 414
 QY 421 WPDLSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLGALIHNNTHL 480
 Db 415 WPDLSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLGALIHNNTHL 474
 QY 481 CFVHTVPMDQLFRNPHQALLHTANRPEDECVEGGLACHQLCARHGWGPGPTQCVNCSQF 540
 Db 475 CFVHTVPMDQLFRNPHQALLHTANRPEDECVEGGLACHQLCARHGWGPGPTQCVNCSQF 534
 QY 541 LRQECVEECRVLOGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 600
 Db 535 LRQECVEECRVLOGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 594
 QY 601 FCVARCPGVKPDLSYMPKFWPDEEGACQPCPNCTHSCVDLDDKGCFAEQRASPLTSI 660
 Db 595 FCVARCPGVKPDLSYMPKFWPDEEGACQPCPNCTHSCVDLDDKGCFAEQRASPLTSI 654
 QY 661 VSAVVGILLVVVLGVVFGILIKRQOKIRKKTMRRLLOETELVEPLTTPSGAMPNQAOMRI 720
 Db 655 VSAVVGILLVVVLGVVFGILIKRQOKIRKKTMRRLLOETELVEPLTTPSGAMPNQAOMRI 714
 QY 721 LKETELRKVKVLGSGAFGTGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM 780
 Db 715 LKETELRKVKVLGSGAFGTGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM 774

| | | | |
|----------|--|---|------|
| QY | 781 | AGVSPYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRRGLSGSODLLNWCQIAKGM | 840 |
| Db | 775 | AGVSPYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRRGLSGSODLLNWCQIAKGM | 834 |
| QY | 841 | YLEDVRLVHRLDAAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKHWALESI | 900 |
| Db | 835 | YLEDVRLVHRLDAAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKHWALESI | 894 |
| QY | 901 | LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYMI | 960 |
| Db | 895 | LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYMI | 954 |
| QY | 961 | MVKCMIDSECRPRELVSERFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLDDDM | 1020 |
| Db | 955 | MVKCMIDSECRPRELVSERFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLDDDM | 1014 |
| QY | 1021 | GDLDVAEEVLPQGGFFCPDPAPGAGGMYHRRSSSTRSGGDLTLGLEPSEEAAPRSP | 1080 |
| Db | 1015 | GDLDVAEEVLPQGGFFCPDPAPGAGGMYHRRSSSTRSGGDLTLGLEPSEEAAPRSP | 1074 |
| QY | 1081 | LAPSEGAGSDVFDGDLGMAKGLQSLPTDPSPLQRYSEDTVPLPSETDGYVAPLTC | 1140 |
| Db | 1075 | LAPSEGAGSDVFDGDLGMAKGLQSLPTDPSPLQRYSEDTVPLPSETDGYVAPLTC | 1134 |
| QY | 1141 | POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPQKNGVWVDVFAFGAVENP | 1200 |
| Db | 1135 | POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPQKNGVWVDVFAFGAVENP | 1194 |
| QY | 1201 | EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVP | 1260 |
| Db | 1195 | EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVP | 1254 |
| QY | 1261 | V 1261 | |
| Db | 1255 | V 1255 | |
| RESULT 2 | | | |
| ID | AAE12130 | standard; Protein; 1255 AA. | |
| AC | AAE12130; | | |
| XX | 18-DEC-2001 | (first entry) | |
| DT | Human tyrosine kinase-type receptor, HER-2. | | |
| DE | | | |
| KW | Therapeutic compound; major histocompatibility complex; vaccine; antigenic peptide; MHC; immunoregulatory; immune response; HER-2; adoptive immunotherapy; anti-cancer; breast cancer antigen; APC; antigen presenting cell; human; tyrosine kinase-type receptor. | | |
| OS | Homo sapiens. | | |
| XX | | | |
| EH | Key | Location/Qualifiers | |
| FT | Region | 774..782 | |
| FT | | /note= "Antigenic epitope" | |
| XX | | | |
| XX | WO200168677-A2. | | |
| XX | 20-SEP-2001. | | |
| XX | | | |
| XX | 16-MAR-2001; 2001WO-US40328. | | |
| XX | | | |
| XX | 16-MAR-2000; 2000US-0527487. | | |
| XX | | | |
| PA | (GENZ) GENZYME CORP. | | |
| PI | Nicolette CA; | | |
| XX | | | |
| XX | WPI; 2001-616284/71. | | |
| DR | N-PSDB; AAD19731. | | |
| XX | | | |

| | | | |
|--|---|--|-----|
| PT | Novel synthetic therapeutic compound for inducing immune response and for use in adoptive immunotherapy, has enhanced binding to major histocompatibility molecules and enhanced immunoregulatory properties | | |
| PT | | | |
| PT | | | |
| XX | | | |
| PS | Claim 4; Page 63-67; 69pp; English. | | |
| XX | | | |
| CC | The invention relates to synthetic therapeutic compounds (antigenic peptides) with enhanced binding to major histocompatibility complex (MHC) molecules and enhanced immunoregulatory properties relative to their natural counterparts. Compounds of the invention are useful for inducing an immune response in a subject and for use in adoptive immunotherapy. They are useful as components of anti-cancer vaccines and to expand immune effector cells that are specific for cancers characterised by expression of the breast cancer antigen, HER-2. | | |
| CC | Polyucleotides that encode peptides of the invention are useful as hybridisation probes and as primers for the detection of genes of gene transcripts that are expressed in antigen presenting cells (APCs), to confirm transduction of polynucleotides into host cells. The present sequence is human tyrosine kinase-type receptor, HER-2. Compounds of the invention are designed based on the HER-2 antigenic peptide (774-782). | | |
| CC | | | |
| XX | | | |
| SQ | Sequence 1255 AA; | | |
| Query Match | | | |
| Best Local Similarity 98.1%; Score 6708; DB 22; Length 1255; | | | |
| Matches 1242; Conservative 2; Mismatches 11; Indels 6; Gaps 1; | | | |
| QY | 1 | MELAAALCRWGLLLALLPFGAASTQVCTGTDKMLRLPASPTHLDMLRLHYQCGVQGNL | 60 |
| Db | 1 | MELAAALCRWGLLLALLPFGAASTQVCTGTDKMLRLPASPTHLDMLRLHYQCGVQGNL | 60 |
| QY | 61 | ELTYLPTNASISFLQDIOEVQGYVLI AHNQVRQVFLRLRIVRGTQFSDNTALAVLDNG | 120 |
| Db | 61 | ELTYLPTNASISFLQDIOEVQGYVLI AHNQVRQVFLRLRIVRGTQFSDNTALAVLDNG | 120 |
| QY | 121 | DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQNPOLCQVDTILWKDIFHKNQOLA | 180 |
| Db | 121 | DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQNPOLCQVDTILWKDIFHKNQOLA | 180 |
| QY | 181 | LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVCAAGCARCKGLPTDCCHEQC | 240 |
| Db | 181 | LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVCAAGCARCKGLPTDCCHEQC | 240 |
| QY | 241 | AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE-----GRYTFGAS | 294 |
| Db | 241 | AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE-----GRYTFGAS | 294 |
| QY | 301 | CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTORCEKCKPCARVCYGLGMEHLREVR | 360 |
| Db | 295 | CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTORCEKCKPCARVCYGLGMEHLREVR | 354 |
| QY | 361 | AVTSANIOEFAGCKKIFGSLAFPLPESPDGDPASNTAPLQPEQLQVFTLEETGYLYISA | 420 |
| Db | 355 | AVTSANIOEFAGCKKIFGSLAFPLPESPDGDPASNTAPLQPEQLQVFTLEETGYLYISA | 414 |
| QY | 421 | WPSLPLDSVFNQNLQVIRGRILHNGAYSILTLQGLISWLGRLSRLSGLALIHNTHL | 480 |
| Db | 415 | WPSLPLDSVFNQNLQVIRGRILHNGAYSILTLQGLISWLGRLSRLSGLALIHNTHL | 474 |
| QY | 481 | CFVHTVPMWDLFRNPHQALLHTANRPEDECYGEGLACHOLCARHCWGPGTQCVCNCSQF | 540 |
| Db | 475 | CFVHTVPMWDLFRNPHQALLHTANRPEDECYGEGLACHOLCARHCWGPGTQCVCNCSQF | 534 |
| QY | 541 | LRGQECVVECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFPGPEAQCVACAHYKQPP | 600 |
| Db | 535 | LRGQECVVECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFPGPEAQCVACAHYKQPP | 594 |
| QY | 601 | FCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCABQASPLTSI | 660 |
| Db | 595 | FCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCABQASPLTSI | 654 |

QY 661 VSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOQMRI 720
Db 655 VSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOQMRI 714
QY 721 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKSILDEAYYM 780
Db 715 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKSILDEAYYM 774
QY 781 AGVGSPPVSRLLGICLTSTVQLVLTOLMPYGCLLDHVRENRRGLGSODLNNWCWQAKGMS 840
Db 775 AGVGSPPVSRLLGICLTSTVQLVLTOLMPYGCLLDHVRENRRGLGSODLNNWCWQAKGMS 834
QY 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESI 894
QY 901 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPQPPCTIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPQPPCTIDVYMI 954
QY 961 MYKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MYKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
QY 1021 GDLVDAEEYLVPOQGFCCPDPAFGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db 1015 GDLVDAEEYLVPOQGFCCPDPAFGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1074
QY 1081 LAPSEGAGSDVFDGLGMAAGKLOSLPHDTHSPQLQYSEDPTVPLPSTDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGLGMAAGKLOSLPHDTHSPQLQYSEDPTVPLPSTDGYVAPLTCS 1134
QY 1141 POPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAGGAVENP 1194
QY 1201 EYLTPQGAAPQPHPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPQGAAPQPHPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
QY 1261 V 1261
Db 1255 V 1255
RESULT 3
ID AAB60167
XX AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
XX AAB60167;
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH) GENENTECH INC.
XX
PI Erickson S, Schwall R;

XX WPI: 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 98.1%; Score 6708; DB 22; Length 1255;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 1242; Conservative 2; Mismatches 11; Indels 6; Gaps 1;
QY 1 MELAALCRWGLLLALLPFGAASQVCTGDMKMLRLPASPEHLDMRLHLYQGCQVVOGNL 60
Db 1 MELAALCRWGLLLALLPFGAASQVCTGDMKMLRLPASPEHLDMRLHLYQGCQVVOGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNNQVQVPLQRLIRVGTQTFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNNQVQVPLQRLIRVGTQTFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELOLRSLTEILKGGVLTORNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPTVGASPGGLRELOLRSLTEILKGGVLTORNPOLCYQDTILWKDIFHKNNOLA 180
QY 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACARCKPLPTDCCHEOC 240
Db 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACARCKPLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICBLHCPALVFNFTVSFWLRVPKVSASHLERYTFGAS 300
Db 241 AAGCTGPKHSDCLACLFHNSGICBLHCPALVFNFTVSFWLRVPKVSASHLERYTFGAS 294
QY 301 CVTACPNYVLTSDVGSCTLVCPHNOEVTAEQGTORCEKSPKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYVLTSDVGSCTLVCPHNOEVTAEQGTORCEKSPKPCARVCYGLGMEHLREVR 354
QY 361 AVTSANIQEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFPETLEEITGYLYISA 420
Db 355 AVTSANIQEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFPETLEEITGYLYISA 414
QY 421 WPDLSPLDSVFONLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLSGLALIHNNTHL 480
Db 415 WPDLSPLDSVFONLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLSGLALIHNNTHL 474
QY 481 CFVHTVPMWDLFRNPHQALLHTANPEDECEVGEGLACHQLCARGHCWGPGTQCNCVSOF 540
Db 475 CFVHTVPMWDLFRNPHQALLHTANPEDECEVGEGLACHQLCARGHCWGPGTQCNCVSOF 534
QY 541 LRQGECEBECEVLRQGLPREYVNNARHCLPCHPCQPNQSGSVTCFGEADOCVACAHYKDP 600
Db 535 LRQGECEBECEVLRQGLPREYVNNARHCLPCHPCQPNQSGSVTCFGEADOCVACAHYKDP 594
QY 601 FCVARCPGKVPDLSYMPIWKPFBEGACQPCPINCTHSCVDLDDKGCQPAEQASPLTSI 660
Db 595 FCVARCPGKVPDLSYMPIWKPFBEGACQPCPINCTHSCVDLDDKGCQPAEQASPLTSI 654
QY 661 VSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOQMRI 720
Db 655 VSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOQMRI 714
QY 721 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKSILDEAYYM 780


```
Db 715 LKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDYAYM 774
QY 781 AGVGSFYVSRLLIGICLTSTVQLVTQMPYGCCLLDHVRENRRGLSGODLLNWCNQIAKGS 840
Db 775 AGVGSFYVSRLLIGICLTSTVQLVTQMPYGCCLLDHVRENRRGLSGODLLNWCNQIAKGS 834
QY 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMMALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMMALESI 894
QY 901 LRRRFTHQSDVMSYGYVTVMELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICITIDVYMI 960
Db 895 LRRRFTHQSDVMSYGYVTVMELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICITIDVYMI 954
QY 961 MYKCMWIDSECRPRRELVSFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MYKCMWIDSECRPRRELVSFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
QY 1021 GDLVDAEYLVPOQGGFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSP 1080
Db 1015 GDLVDAEYLVPOQGGFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSP 1074
QY 1081 LAPSEGAGSDVFDGDLGMAAGLQSLPHTHPSPLQRYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAGLQSLPHTHPSPLQRYSEDPVPLPSETDGYVAPLTCS 1134
QY 1141 POPEYVNOQDVRPQPPSPREGPLPAARPAGATILERAKTLSPGKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNOQDVRPQPPSPREGPLPAARPAGATILERAKTLSPGKNGVVKDVFAGGAVENP 1194
QY 1201 EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
QY 1261 V 1261
Db 1255 V 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-APR-2002 (first entry)
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW stomal disorder; macrophagal disorder; epithelial disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
FN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX
```

```
PI Erickson S, Schwall R, Sliwkowski M;
XX WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
```

```
Query Match 98.1%; Score 6708; DB 23; Length 1255;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 1242; Conservative 2; Mismatches 11; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPGGAASQVCTGTDMLRLPASPEHLDMRLHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPGGAASQVCTGTDMLRLPASPEHLDMRLHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQGYVLIHNVQVPLQLRLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIOEVQGYVLIHNVQVPLQLRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLTQRPOLCVQDTILWKDIPKKNQLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLTQRPOLCVQDTILWKDIPKKNQLA 180
QY 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACRCKPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACRCKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVFNNFTVSFWLRVPKVSASHLERYTFGAS 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVFNTDTFESMPNPE-----GRYTFGAS 294
QY 301 CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTORCEKSPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTORCEKSPCARVCYGLGMEHLREVR 354
QY 361 AVTSANIOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEOLQVFETLEEITGYLYISA 420
Db 355 AVTSANIOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEOLQVFETLEEITGYLYISA 414
QY 421 WDSLPLDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGSLRSLRELSSGLALIHNTHL 480
Db 415 WDSLPLDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGSLRSLRELSSGLALIHNTHL 474
QY 481 CFVHTVPMWDLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQF 540
Db 475 CFVHTVPMWDLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQF 534
QY 541 LRGOECVVEECRVLOGLPREYVYNARHCLPCHPECPONGSVTCFGEADOCVACAHYKOPP 600
Db 535 LRGOECVVEECRVLOGLPREYVYNARHCLPCHPECPONGSVTCFGEADOCVACAHYKOPP 594
QY 601 FCVARCFSGVKPDLSPYMPIMWKFPEBEGACQPCPINCTHSCVDLDDKGCFAEORASPLTSI 660
```

| | | | |
|----------|---|--|---------------------|
| Db | 595 | FCVACPSGVKPOLSYMPIWKPFDEBEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTSI | |
| Qy | 661 | VSANVGILLVWLVGVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQOMRI | |
| Db | 655 | VSANVGILLVWLVGVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQOMRI | |
| Qy | 721 | LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVM | |
| Db | 715 | LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVM | |
| Qy | 781 | AGVGSYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDLLNWCQIAKMS | |
| Db | 775 | AGVGSYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDLLNWCQIAKMS | |
| Qy | 841 | YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMMALESI | |
| Db | 835 | YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMMALESI | |
| Qy | 901 | LRRRFTHQSDVMSGYVTVMELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYMI | |
| Db | 895 | LRRRFTHQSDVMSGYVTVMELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYMI | |
| Qy | 961 | MVKCMWIDSECRPRFRELVSFERSMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM | |
| Db | 955 | MVKCMWIDSECRPRFRELVSFERSMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM | |
| Qy | 1021 | GOLVDAEYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP | |
| Db | 1015 | GOLVDAEYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP | |
| Qy | 1081 | LAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSBTGCVAPLTC | |
| Db | 1075 | LAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSBTGCVAPLTC | |
| Qy | 1141 | POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVPFPGGAVENP | |
| Db | 1135 | POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVPFPGGAVENP | |
| Qy | 1201 | EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAAPPSTFKGPTAENPEYLGLDVP | |
| Db | 1195 | EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAAPPSTFKGPTAENPEYLGLDVP | |
| Qy | 1261 | V 1261 | |
| Db | 1255 | V 1255 | |
| RESULT 5 | | | |
| AAW01111 | ID AAW01111 standard; Protein; 1255 AA. | | |
| XX | AC | AAW01111; | |
| XX | DT | 01-JAN-1997 (first entry) | |
| XX | DE | HER-2/new protein. | |
| KW | KW | HER-2/new; c-erbB2; p185; oncogene; tyrosine protein kinase; breast cancer; ovary cancer; colon cancer; lung cancer; prostate cancer; immunisation; tumour; vaccine; vector. | |
| OS | OS | Homo sapiens. | |
| FT | FT | Key | Location/Qualifiers |
| FT | FT | Domain | 676..1255 |
| XX | XX | /label= Intracellular domain | |
| XX | XX | /note= "claimed domain, useful for immunisation" | |
| PN | PN | WO9630514-A1. | |
| PD | PD | 03-OCT-1996. | |
| XX | XX | | |

| | | | |
|----|---|---|-----|
| PF | 28-MAR-1996; | 96WO-USO1689. | |
| XX | 31-MAR-1995; | 95US-0414417. | |
| XX | (UNIW) UNIV WASHINGTON. | | |
| XX | Cheever MA, Disis ML; | | |
| XX | WPI: 1996-455361/45. | | |
| DR | N-PSDB; AAT40739. | | |
| XX | DNA encoding HER-2-neu poly:peptide(s) - used for prevention or treatment of malignancies with which the HER-2/neu oncogene is associated | | |
| PS | Claim 2; Page 56-61; 71pp; English. | | |
| XX | Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is the product of the HER-2/neu oncogene (see also AAT40739). The protein is over-expressed in various cancers, including breast, ovarian, colon, lung and prostate. The intracellular domain of the protein can be used to immunise an animal against a malignancy with which the oncogene is associated. The polypeptide can be produced in transformed host cells for use in immunisation. Alternatively, animal cells are transfected in vivo or ex vivo with a viral vector that directs expression of the polypeptide. | | |
| XX | Sequence 1255 AA; | | |
| Qy | Query Match | 98.0%; Score 6702; DB 17; Length 1255; | |
| Db | Best Local Similarity | 98.3%; Pred. No. 0; | |
| XX | Matches 1240; Conservative | 3; Mismatches 12; Indels 6; Gaps 1; | |
| Qy | 1 | MELAAALCRWGLLLALLPGCAASTQVCTGDMKRLRIPASPTHLDMLRHLYQGCQVVGQNL | 60 |
| Db | 1 | MELAAALCRWGLLLALLPGCAASTQVCTGDMKRLRIPASPTHLDMLRHLYQGCQVVGQNL | 60 |
| Qy | 61 | ELTYLPTNASLSFLODIOEVQGYVLI AHNVQVPLQRLRI VRGTOLFFDNALAVLDNG | 120 |
| Db | 61 | ELTYLPTNASLSFLODIOEVQGYVLI AHNVQVPLQRLRI VRGTOLFFDNALAVLDNG | 120 |
| Qy | 121 | DPLNNTTPTVGASPGGLRELQRLSRLTEILKGGVLIQRNPQLCYQDITLWKDIFPHKNQLA | 180 |
| Db | 121 | DPLNNTTPTVGASPGGLRELQRLSRLTEILKGGVLIQRNPQLCYQDITLWKDIFPHKNQLA | 180 |
| Qy | 181 | LTLIDTNRSRACHPCSPMKSGRCGSESDCQSLTRTVACGCCARCKGPLPTDCCHEQC | 240 |
| Db | 181 | LTLIDTNRSRACHPCSPMKSGRCGSESDCQSLTRTVACGCCARCKGPLPTDCCHEQC | 240 |
| Qy | 241 | AAGCTGPKHSDCLACLFHNSGICELHCPALVFNNFTVSFWLRVPKVSASHLERYTFGAS | 300 |
| Db | 241 | AAGCTGPKHSDCLACLFHNSGICELHCPALVFNNFTVSFWLRVPKVSASHLERYTFGAS | 294 |
| Qy | 301 | CVTACPYNYLSTDVGSCTLVCPPLHNQEVTAEDGTORCEKCSKPCARVCYGLGMEHLREVR | 360 |
| Db | 295 | CVTACPYNYLSTDVGSCTLVCPPLHNQEVTAEDGTORCEKCSKPCARVCYGLGMEHLREVR | 354 |
| Qy | 361 | AVTSANIQEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPBOLOVPETLEETGYLYISA | 420 |
| Db | 355 | AVTSANIQEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPBOLOVPETLEETGYLYISA | 414 |
| Qy | 421 | WPDLSPLDLSVFONLQVIRGRILHNGAYSILTLQGLGSIWGLRSLRGLSGGLAIHHNTHL | 480 |
| Db | 415 | WPDLSPLDLSVFONLQVIRGRILHNGAYSILTLQGLGSIWGLRSLRGLSGGLAIHHNTHL | 474 |
| Qy | 481 | CFVHTVPMDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCWGPGTQCVCNCSQF | 540 |
| Db | 475 | CFVHTVPMDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCWGPGTQCVCNCSQF | 534 |
| Qy | 541 | LRGQECVBECEVRLQGLPREYVNAHCLPCHPCQFQNGSVTCFGEADOCVACAHYKDP | 600 |
| Db | 535 | LRGQECVBECEVRLQGLPREYVNAHCLPCHPCQFQNGSVTCFGEADOCVACAHYKDP | 594 |

| | | | |
|----------|---|---|------|
| Qy | 601 | FCVARCPSGVKPDLSYMIKFPDEEGACQPCPINCTHSCVDLDDKGCPCAEORASPLTSI | 660 |
| Ds | 595 | FCVARCPSGVKPDLSYMIKFPDEEGACQPCPINCTHSCVDLDDKGCPCAEORASPLTSI | 654 |
| Qy | 661 | VSADVGLLVVGVVGVGILLIKRQOKIRKVTMRRLQETELVEPLTPSGAMPNQAOQRI | 720 |
| Ds | 655 | ISADVGLLVVGVVGVGILLIKRQOKIRKVTMRRLQETELVEPLTPSGAMPNQAOQRI | 714 |
| Qy | 721 | LKETELRKVKVLSGAFGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM | 780 |
| Ds | 715 | LKETELRKVKVLSGAFGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM | 774 |
| Qy | 781 | AGVSPYVSRLLGLCLSTVOLVTOLMPYGCGLLDHVRNRRGLSGQDILLNMCQIAKMS | 840 |
| Ds | 775 | AGVSPYVSRLLGLCLSTVOLVTOLMPYGCGLLDHVRNRRGLSGQDILLNMCQIAKMS | 834 |
| Qy | 841 | YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESI | 900 |
| Ds | 835 | YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESI | 894 |
| Qy | 901 | LRRRFTHQSDVWSYGVTWELMTFGAKPYDGI PAREIPDLLLEKGERLPQPPICTIDVYMI | 960 |
| Ds | 895 | LRRRFTHQSDVWSYGVTWELMTFGAKPYDGI PAREIPDLLLEKGERLPQPPICTIDVYMI | 954 |
| Qy | 961 | MVKCWMIDSECRPRELIVSEFSRMARDPQRVFIQNEIDLGPASPLDSTFYRSLLEDDM | 1020 |
| Ds | 955 | MVKCWMIDSECRPRELIVSEFSRMARDPQRVFIQNEIDLGPASPLDSTFYRSLLEDDM | 1014 |
| Qy | 1021 | GDLVDAEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP | 1080 |
| Ds | 1015 | GDLVDAEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP | 1074 |
| Qy | 1081 | LAPSEGAGSVDGDLGGAAGKQLSLPTHPDPSPLQRYSEDPTVPLPSETDGYVAPLTCS | 1140 |
| Ds | 1075 | LAPSEGAGSVDGDLGGAAGKQLSLPTHPDPSPLQRYSEDPTVPLPSETDGYVAPLTCS | 1134 |
| Qy | 1141 | POPEYVNOPVRPOPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVPFAFGAVENP | 1200 |
| Ds | 1135 | POPEYVNOPVRPOPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVPFAFGAVENP | 1194 |
| Qy | 1201 | EYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP | 1260 |
| Ds | 1195 | EYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP | 1254 |
| Qy | 1261 | V 1261 | |
| Ds | 1255 | V 1255 | |
| RESULT 6 | | | |
| ID | AAW92406 | AAW92406 standard; Protein; 1255 AA. | |
| AC | AAW92406; | | |
| DT | 21-APR-1999 | (first entry) | |
| XX | Human HER-2/neu oncogene protein. | | |
| DE | HER-2/neu; oncogene; immune response; T cell; B cell; immunisation; | | |
| KW | malignancy; treatment; tumour. | | |
| XX | Homo sapiens. | | |
| XX | Key | Location/Qualifiers | |
| FT | Region | 676..1255 | |
| FT | /note= "region which elicits immune response" | | |
| XX | US5869445-A. | | |
| PN | 09-FEB-1999. | | |
| XX | 01-APR-1996; | 96US-0625101. | |
| PF | | | |

| | | | |
|----|---|---|-----|
| XX | 01-APR-1996; | 96US-0625101. | |
| PR | 17-MAR-1993; | 93US-0033644. | |
| PR | 12-AUG-1993; | 93US-0106112. | |
| XX | 31-MAR-1995; | 95US-0414417. | |
| PA | (UNIM) UNIV WASHINGTON. | | |
| XX | Cheever MA, Disis ML; | | |
| XX | WPI; 1999-152835/13. | | |
| DR | N-FSDB; AAX01912. | | |
| XX | Use of HER-2/neu polypeptides - for eliciting an immune response to | | |
| PT | an HER-2/neu associated malignancy, particularly for treating or | | |
| PT | preventing tumours | | |
| XX | Claim 3; Column 31-38; 26pp; English. | | |
| PS | This sequence represents the human HER-2/neu oncogene protein. A fragment | | |
| CC | of this protein is used in a method for eliciting or enhancing an immune | | |
| CC | response to HER-2/neu protein. The polypeptide can stimulate T cells and | | |
| CC | B cells to produce an immune response to the HER-2/neu protein. The | | |
| CC | method can be used for immunisation against a malignancy in which the | | |
| CC | HER-2/neu oncogene is associated and in the treatment of an existing | | |
| CC | tumour, or to prevent tumour occurrence or reoccurrence. | | |
| XX | Sequence 1255 AA; | | |
| SQ | Query Match 98.0%; Score 6702; DB 20; Length 1255; | | |
| | Best Local Similarity 98.3%; Pred. No. 0; | | |
| | Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1; | | |
| Qy | 1 | MELAALCRWGLLLALLPPGAASCTCTGDMKRLPASPTHLMRLHYGQCVVQGNL | 60 |
| Ds | 1 | MELAALCRWGLLLALLPPGAASCTCTGDMKRLPASPTHLMRLHYGQCVVQGNL | 60 |
| Qy | 61 | ELTYLPTNASLSFLQDIQEVGYVLI AHNQVRQVPLQRLRIVRGTLQFEDNVYALVDNG | 120 |
| Ds | 61 | ELTYLPTNASLSFLQDIQEVGYVLI AHNQVRQVPLQRLRIVRGTLQFEDNVYALVDNG | 120 |
| Qy | 121 | DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA | 180 |
| Ds | 121 | DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA | 180 |
| Qy | 181 | LTLLDITNRSRACHPCSPMCKGSRGCMGSESSDCQSLTRTVCAAGCARCKGPLTDCCHQC | 240 |
| Ds | 181 | LTLLDITNRSRACHPCSPMCKGSRGCMGSESSDCQSLTRTVCAAGCARCKGPLTDCCHQC | 240 |
| Qy | 241 | AAGCTGPKHSDCLACLFHNSGICELHCPALVFNNFTVSWFLRVPKVSASHLEYRTFGAS | 300 |
| Ds | 241 | AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE-----GRYTFGAS | 294 |
| Qy | 301 | CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR | 360 |
| Ds | 295 | CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR | 354 |
| Qy | 361 | AVTSANTIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVFETLEITGYLYISA | 420 |
| Ds | 355 | AVTSANTIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVFETLEITGYLYISA | 414 |
| Qy | 421 | WPDLSPLDSLVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSURELSGLALIHNNTHL | 480 |
| Ds | 415 | WPDLSPLDSLVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSURELSGLALIHNNTHL | 474 |
| Qy | 481 | CFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAGLCHQLCARGCHGPGPTQCVNCSQF | 540 |
| Ds | 475 | CFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAGLCHQLCARGCHGPGPTQCVNCSQF | 534 |
| Qy | 541 | LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCTGPEADQCVACAHYKDDP | 600 |
| Ds | 535 | LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCTGPEADQCVACAHYKDDP | 594 |

Qy 601 FCVACPSGVKPDLSYMPKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSI 660
Db 595 FCVACPSGVKPDLSYMPKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSI 654
Qy 661 VSAVVGILLVVLGVVGFILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNOAQMRI 720
Db 655 ISAVVGILLVVLGVVGFILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNOAQMRI 714
Qy 721 LKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILLDEAYVM 780
Db 715 LKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILLDEAYVM 774
Qy 781 AGVSPYVSRLLGLCLTSTVOLVQLMPYGCGLDHRNRLGSGQDLNMCQIAKMS 840
Db 775 AGVSPYVSRLLGLCLTSTVOLVQLMPYGCGLDHRNRLGSGQDLNMCQIAKMS 834
Qy 841 YLEVDRLVHRDLAARNLVKSPNVKVTDFGLARLLDIDETEHADGGKVPDKWMALESI 900
Db 835 YLEVDRLVHRDLAARNLVKSPNVKVTDFGLARLLDIDETEHADGGKVPDKWMALESI 894
Qy 901 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFSEFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFSEFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDM 1014
Qy 1021 GDLVDAEYLVPOQGFPCDPAPAGAGVHHRHSSSTRSGGDLITLGLPSEBEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFPCDPAPAGAGVHHRHSSSTRSGGDLITLGLPSEBEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERPKTLPSPKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPQCGAARQPHPPAFSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPQCGAARQPHPPAFSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255
RESULT 7
ID AAB21198 standard; protein; 1255 AA.
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX
OS Homo sapiens.
XX
PN WO200044899-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX

(CORI-) CORIXA CORP.
PA (SMIK) SMITHLINE BEECHAM.
XX
PI Cheever MA, Gheysen D;
XX
DR WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
PS Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6702; DB 21; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYGQCQVVOGNI 60
Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYGQCQVVOGNI 60
Qy 61 ELTYLPTNASLSFLQDIQEVQYVLI AHNQVRQVPLRLRIVRGTLQFEDNVALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQYVLI AHNQVRQVPLRLRIVRGTLQFEDNVALAVLNG 120
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLRILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLRILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVAGGCARCKGPLTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVAGGCARCKGPLTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNFTVSFWLRVPKVSASHLRYTFGAS 300
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNFTVSFWLRVPKVSASHLRYTFGAS 294
Qy 301 CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANTQEFAGCKKIFGSLAFLEPESFDGDPASNTAPLQPEQLQVFETLEITGYLYISA 420
Db 355 AVTSANTQEFAGCKKIFGSLAFLEPESFDGDPASNTAPLQPEQLQVFETLEITGYLYISA 414
Qy 421 WPDSLPLDSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELGGSLALHNNTHL 480
Db 415 WPDSLPLDSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELGGSLALHNNTHL 474
Qy 481 CFVHTVPWDQLFRNPHOALLHTANRPEBCEVGEGLACHQLCARGHGWPGTQCVCNCSQF 540
Db 475 CFVHTVPWDQLFRNPHOALLHTANRPEBCEVGEGLACHQLCARGHGWPGTQCVCNCSQF 534
Qy 541 LRQGECEVCECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 600
Db 535 LRQGECEVCECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 594
Qy 601 FCVACPSGVKPDLSYMPKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSI 660

```
Db 595 FCVARCPGKVPDLSTYMPIMKFPDEEGACOPCPINCTHSCVDLDDKGCAPQASPLTSI 654
Qy 661 VSAVVGILLVVLGVVFGIILKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAWRI 720
Db 655 ISAVVGILLVVLGVVFGIILKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAWRI 714
Qy 721 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVM 780
Db 715 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVM 774
Qy 781 AGVSPVYVRLGLICLTSTVQLTQMLPYGCLLDHVRENRGRGSLDQLLNWCQIAKMS 840
Db 775 AGVSPVYVRLGLICLTSTVQLTQMLPYGCLLDHVRENRGRGSLDQLLNWCQIAKMS 834
Qy 841 YLEDRVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHYHAGDGKVPKIMWALESI 900
Db 835 YLEDRVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHYHAGDGKVPKIMWALESI 894
Qy 901 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFSEFMRDPORFVVIQNEGLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFSEFMRDPORFVVIQNEGLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOGFFCPDPAPCAGGMVHRHRSSTRSGGDLTLGLEPSEEEARSP 1080
Db 1015 GDLVDAEYLVPOGFFCPDPAPCAGGMVHRHRSSTRSGGDLTLGLEPSEEEARSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGMAAKGLQSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAKGLQSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1134
Qy 1141 POEYVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENP 1200
Db 1135 POEYVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLGIDVP 1260
Db 1195 EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLGIDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 8
AAV84780
ID AAY84780 standard; Protein; 1255 AA.
XX
AC AAY84780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbb-2 receptor protein.
XX
KW SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX
OS Homo sapiens.
XX
PN W0200020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
```

```
(UYMC-) UNIV MCMASTER.
XX
PI Muller WJ, Siegel PM;
XX
DR WPI; 2000-303768/26.
XX N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
XX erbb-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPLICE erbb-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbb-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbb-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbb-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbb-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6702; DB 21; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPETHLDMRLHLYGQCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPETHLDMRLHLYGQCQVQGNL 60
Qy 61 ELTYLPTNASLFLQDIEQVQVYLIAHNOVRQVPLQRLRIVRGTLQFEDNYALVLDNG 120
Db 61 ELTYLPTNASLFLQDIEQVQVYLIAHNOVRQVPLQRLRIVRGTLQFEDNYALVLDNG 120
Qy 121 DPLNTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLDITNRSRACHPCSPMKGSCRCWGESSEDCQSLTRTVCGAGCARCKGPLTDCCHQC 240
Db 181 LTLDITNRSRACHPCSPMKGSCRCWGESSEDCQSLTRTVCGAGCARCKGPLTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVFNNFTVSEFWLRVPKVSASHLEYTFGAS 300
Db 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFFSMNPPE-----GRYTFGAS 294
Qy 301 CVTACPNYLSLTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANTQEPAGCKKIFGSLAFIPESPDGPASNTAPLOPEQLQVFTLEBITGYLYISA 420
Db 355 AVTSANTQEPAGCKKIFGSLAFIPESPDGPASNTAPLOPEQLQVFTLEBITGYLYISA 414
Qy 421 WPDSLPLDSVFNQLQVIRGRILHNGAYSLLTQGLGISWLGSLRSRELGSGLALIHNNTHL 480
Db 415 WPDSLPLDSVFNQLQVIRGRILHNGAYSLLTQGLGISWLGSLRSRELGSGLALIHNNTHL 474
Qy 481 CFVHTVPWDLFRNPHQALLHTANRPEDECVGEGLACHQLCARHGCPGPTQCVCNCSQF 540
Db 475 CFVHTVPWDLFRNPHQALLHTANRPEDECVGEGLACHQLCARHGCPGPTQCVCNCSQF 534
Qy 541 LRQGECEVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCTFGPEADQCVCACAHYKDDP 600
Db 535 LRQGECEVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCTFGPEADQCVCACAHYKDDP 594
Qy 601 FCVARCPGKVPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSI 660
```



```
Db 655 ISAVVGVLLVVVLGVGVGILIKRQKIRKYMRRLLQETELVEPLTPSGAMPNOAQMRI 714
Qy 721 LKETELRKVKVLSGAFGTGVYGIWIPDGENVKIPVAIKVLRENTSPPKANKEILDEAYVM 780
Db 715 LKETELRKVKVLSGAFGTGVYGIWIPDGENVKIPVAIKVLRENTSPPKANKEILDEAYVM 774
Qy 781 AGVCSPVVSRLLGICLTSTVOLVTLMPYGCGLLDHVRENRRGLSGDQLLNKCMQIAKMS 840
Db 775 AGVSPVVSRLGICLTSTVOLVTLMPYGCGLLDHVRENRRGLSGDQLLNKCMQIAKMS 834
Qy 841 YLEVDRLVHRDLAARNLVKSPNHNKIDTFGLARLLDIDETEHADGGKVPKWMALESI 900
Db 835 YLEVDRLVHRDLAARNLVKSPNHNKIDTFGLARLLDIDETEHADGGKVPKWMALESI 894
Qy 901 LRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFCTPDPAPCAGGMVHRRHSSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFCTPDPAPCAGGMVHRRHSSSTRSGGDLTLGLEPSEEEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGWAAGKQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGDLGWAAGKQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNPQDVRPOPPSPREGPLPAARPAAGATLERAKTLSPGKGVVVDVFAFGAVENP 1200
Db 1135 POPEYVNPQDVRPOPPSPREGPLPAARPAAGATLERAKTLSPGKGVVVDVFAFGAVENP 1194
Qy 1201 EYLTPQGAAPQPPHPPAFPAFNLVYWDODPPERGAPPSTFKGTPTAENPEYVLGLDVP 1260
Db 1195 EYLTPQGAAPQPPHPPAFPAFNLVYWDODPPERGAPPSTFKGTPTAENPEYVLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 10
AAG88267
ID AAG88267 standard; Protein: 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
PN W0200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33591.
XX
PR 10-DEC-1999; 99US-0458299.
XX
PA (EPIM-) EPIMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI; 2001-374995/39.
```

```
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX Disclosure; Page 15; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 1255 AA;
```

```
Query Match 98.0%; Score 6702; DB 22; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
```

```
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYGCCVVOGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYGCCVVOGNL 60
Qy 61 ELYLPTNASLFLQDIEQVGVYLIAHNOVRQVPLORLIRVRGTQLPEDNVALLVDNG 120
Db 61 ELYLPTNASLFLQDIEQVGVYLIAHNOVRQVPLORLIRVRGTQLPEDNVALLVDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNQLA 180
Qy 181 LTLIDTNRSPACHPCSPMKSGRCWGESSEDCQSLTRTVCGAGCARCKGPLPTDCCHQC 240
Db 181 LTLIDTNRSPACHPCSPMKSGRCWGESSEDCQSLTRTVCGAGCARCKGPLPTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVFNNFTVSFWLRVPKVSASHLRYTFCAS 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTNTDTFESMPNPE-----GRYTFCAS 294
Qy 301 CVTACPNYLSTDVGSCTLVCPLNHQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSTDVGSCTLVCPLNHQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANIQEFAGCKIFGSLAFIPESPDGPASNTAPLQPEQLQVFPETLEITGYLYISA 420
Db 355 AVTSANIQEFAGCKIFGSLAFIPESPDGPASNTAPLQPEQLQVFPETLEITGYLYISA 414
Qy 421 WPDLSPLDSVFNQLQVIRGRILHNGAYSLLTQGLGISWLGISRLSRLGSLALHNNHNL 480
Db 415 WPDLSPLDSVFNQLQVIRGRILHNGAYSLLTQGLGISWLGISRLSRLGSLALHNNHNL 474
Qy 481 CFVHTVPWDLFRNPHQALLHTANRPEDECVGELACHQLCARGHCGMPGTQCVCNSQF 540
|||||
```

Db 475 CFVHTVPMDQLFRNPHOALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVCNSQF 534
Qy 541 LRQGECEECRVLOGLPREYNARHCLPCHPECPONGSVTCFGEADQCACAHYKDDP 600
Db 535 LRQGECEECRVLOGLPREYNARHCLPCHPECPONGSVTCFGEADQCACAHYKDDP 594
Qy 601 FCVARCPGKPDLSYMPIMKFPDDEGACQPCINCHSCVDLDDKGCPCAEQASPLTSI 660
Db 595 FCVARCPGKPDLSYMPIMKFPDDEGACQPCINCHSCVDLDDKGCPCAEQASPLTSI 654
Qy 661 VSAVVGILLVVVLGVVFGILLKROOKIRKKTMRLLQETELVEPLTSGAMPNQAQRI 720
Db 655 ISAVVGILLVVVLGVVFGILLKROOKIRKKTMRLLQETELVEPLTSGAMPNQAQRI 714
Qy 721 LKETELRKVKVLSGAFVTVKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVM 780
Db 715 LKETELRKVKVLSGAFVTVKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVM 774
Qy 781 AGVSPYVSRLLGLCTSTVOLVQLMPYGCLLDHVRENRLGSDQLLNMCQIAKMS 840
Db 775 AGVSPYVSRLLGLCTSTVOLVQLMPYGCLLDHVRENRLGSDQLLNMCQIAKMS 834
Qy 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALESI 894
Qy 901 LRRRFTHOSDVMSYGVTVWELMTFCAPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHOSDVMSYGVTVWELMTFCAPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMIDSECRPRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDM 1020
Db 955 MVKCMIDSECRPRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDM 1014
Qy 1021 GDLVDAEYLYPQOGFCPPDAPGAGGVHRRSSSTRSCGGDLTLGLEPSEEAAPRSP 1080
Db 1015 GDLVDAEYLYPQOGFCPPDAPGAGGVHRRSSSTRSCGGDLTLGLEPSEEAAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGAAKQLSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTC 1140
Db 1075 LAPSEGAGSDVFDGDLGAAKQLSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTC 1134
Qy 1141 POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPKGKGVKDVFAFGAVENP 1200
Db 1135 POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERPKTLPKGKGVKDVFAFGAVENP 1194
Qy 1201 EYLTPOGGAAPQHPHPPAFSFAFNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPOGGAAPQHPHPPAFSFAFNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 11
AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX
AC AAE24067;
XX
DT 23-SEP-2002 (first entry)
DE Human Her-2 protein.
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
OS Homo sapiens.
XX
XX W0200222636-A1.
PN
XX

PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowseert LM;
XX
DR WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
XX inflammation or to prevent infection in humans -
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antisense compounds targeted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6702; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRASPETHLDMRLHLYOCQVVOGQNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRASPETHLDMRLHLYOCQVVOGQNL 60
Qy 61 ELTYLPNNSLFIQDIOEVGYVLI AHNQVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db 61 ELTYLPNNSLFIQDIOEVGYVLI AHNQVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSGRCWGESSEDCQSLTRTVACAGGCARCKGPLTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSGRCWGESSEDCQSLTRTVACAGGCARCKGPLTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNNFTVSPFMRVPKVSASHLEYTFGAS 300
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNNFTVSPFMRVPKVSASHLEYTFGAS 294
Qy 301 CVTACPNYLSLTDVGSCTLVCPHNOVETADGTORCEKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPHNOVETADGTORCEKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANTIOEPAGCKKIFGSLAFIPESFDGPPASNTAPLQEPOLQVFTLEBITGYLYISA 420
Db 355 AVTSANTIOEPAGCKKIFGSLAFIPESFDGPPASNTAPLQEPOLQVFTLEBITGYLYISA 414
Qy 421 WPDSLPLDSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGRLSRLGSLALIHNNTHL 480
Db 415 WPDSLPLDSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGRLSRLGSLALIHNNTHL 474
Qy 481 CFVHTVPMDQLFRNPHOALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVCNSQF 540
Db 475 CFVHTVPMDQLFRNPHOALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVCNSQF 534
Qy 541 LRQGECEECRVLOGLPREYNARHCLPCHPECPONGSVTCFGEADQCACAHYKDDP 600

Db 535 LRQGECECKVLQGLPREYNARHCUPCHPECPQNGSVTCFGEADQCACAHYKDP 594
Qy 601 FCVACRSPGVKPDLSYMPINWKFDPDEGACQPCINCTHSCVDLDDKGCAPQASPLTIS 660
Db 595 FCVACRSPGVKPDLSYMPINWKFDPDEGACQPCINCTHSCVDLDDKGCAPQASPLTIS 654
Qy 661 VSAVVGILLVVVLGVVFGILLIKRQOKIRKYTMRLLOETELVPLTPSGAMPNOAMRI 720
Db 655 ISAVVGILLVVVLGVVFGILLIKRQOKIRKYTMRLLOETELVPLTPSGAMPNOAMRI 714
Qy 721 LKETELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIIDEAYVM 780
Db 715 LKETELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIIDEAYVM 774
Qy 781 AGVSPYVSRLLGLCLTSTVOLVQLMPYGCGLLDHVRNRCGLSGQDLLNWCMIAGMS 840
Db 775 AGVSPYVSRLLGLCLTSTVOLVQLMPYGCGLLDHVRNRCGLSGQDLLNWCMIAGMS 834
Qy 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESI 894
Qy 901 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMWIDSECRPRELVSEFSRMDPQRFVVIQNEGLGASPLDSTFYRSLLDDDD 1020
Db 955 MVKCMWIDSECRPRELVSEFSRMDPQRFVVIQNEGLGASPLDSTFYRSLLDDDD 1014
Qy 1021 GDLVDAEYLVPOQGFCDPAPGAGGWHHRSSSTRSGGGLTILGLEPSEEEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFCDPAPGAGGWHHRSSSTRSGGGLTILGLEPSEEEAPRSP 1074
Qy 1081 LAPSEAGSDVFDGDLGWAAGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEAGSDVFDGDLGWAAGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTLPFGKNGVVKDVFAFGAVENP 1200
Db 1135 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTLPFGKNGVVKDVFAFGAVENP 1194
Qy 1201 EYLTPQGAAPQHPHPPAFSPAFNLVYWDODPPERGAPSPFTKGTPTASNPEYLGLDVP 1260
Db 1195 EYLTPQGAAPQHPHPPAFSPAFNLVYWDODPPERGAPSPFTKGTPTASNPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AC AAE20479;
XX DT 01-JUL-2002 (first entry)
XX DE Human Her-2/neu protein.
XX KW Human, Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX OS Homo sapiens.
XX FH Key
XX FT Region
XX FT 1021..1030
XX PN /note= "Naturally processed HLA-B44-restricted epitope"
XX PN WO200214503-A2.

PD 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
PF 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-234428P.
PR 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
PI Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedwick TS;
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer -
PT Disclosure; Page 114-117; 129pp; English.
PS The invention relates to an isolated Her-2/Neu polypeptide composition
XX effective for eliciting an immune response. The invention is useful for
XX eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
XX present sequence is human Her-2/neu protein.
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6702; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHDLMLRHLVYGCQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHDLMLRHLVYGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEQVGVYLIQHNQVQVPLQRLIRVGRQTQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEQVGVYLIQHNQVQVPLQRLIRVGRQTQFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQHNQVQVPLQRLIRVGRQTQFEDNYALAVLDNG 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQHNQVQVPLQRLIRVGRQTQFEDNYALAVLDNG 180
Qy 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARGCKPLPTDCCHEQC 240
Db 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARGCKPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVFNNFTVFWLRVPKVASHLERYTGAS 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVFNNFTVFWLRVPKVASHLERYTGAS 294
Qy 301 CVTACPNYLSLTDVGSCTLVCPHNSQVTAEDGTQRCCKSKPCARVCYGLGMHLREVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPHNSQVTAEDGTQRCCKSKPCARVCYGLGMHLREVR 354
Qy 361 AVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEITGELYISA 420

```
Db 355 AVTSANIQEPAGCKKIFGSLAFAPESPDGDPASNTAPLQPEQLQVFTLEITGELYISA 414
Qy 421 WPDSLPLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRLREGLGSLALHHNTHL 480
Db 415 WPDSLPLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRLREGLGSLALHHNTHL 474
Qy 481 CFVHTVPMWDLFRNPHQALLHTANRPEDECVEGEGACHQLCARGHGWGPGPTQCVNCSQF 540
Db 475 CFVHTVPMWDLFRNPHQALLHTANRPEDECVEGEGACHQLCARGHGWGPGPTQCVNCSQF 534
Qy 541 LRQGECEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACHYKDDP 600
Db 535 LRQGECEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACHYKDDP 594
Qy 601 FCVARCPGKVPDLSYMPIWKFPDEGACQPCPINCTSHSCVDLDDKCPAEQASPLTISI 660
Db 595 FCVARCPGKVPDLSYMPIWKFPDEGACQPCPINCTSHSCVDLDDKCPAEQASPLTISI 654
Qy 661 VSAVVGILLVVVLGVVFGIILIKRQKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRI 720
Db 655 ISAVVGILLVVVLGVVFGIILIKRQKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRI 714
Qy 721 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 780
Db 715 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 774
Qy 781 AGVGSPPVSRLLGICLTSTVQLMPYGCGLDVRNRLGSGDQLLNWCMQIAKMS 840
Db 775 AGVGSPPVSRLLGICLTSTVQLMPYGCGLDVRNRLGSGDQLLNWCMQIAKMS 834
Qy 841 YLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKGKVPKMWALESI 900
Db 835 YLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKGKVPKMWALESI 894
Qy 901 LRRRFTHOSDWSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDYMI 960
Db 895 LRRRFTHOSDWSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDYMI 954
Qy 961 MVKCMWIDSECRPFRELVSFEFSRMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDDM 1020
Db 955 MVKCMWIDSECRPFRELVSFEFSRMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDDM 1014
Qy 1021 GDLVDAEYLVPOQFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLPSEEEARSP 1080
Db 1015 GDLVDAEYLVPOQFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLPSEEEARSP 1074
Qy 1081 LAPSEGAGSDVFDGLGMAKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGLGMAKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCS 1134
Qy 1141 PQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAFGGAVENP 1200
Db 1135 PQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAFGGAVENP 1194
Qy 1201 EYLTPQGAAPQPHPPAFSAFNLYWDDPPERGAPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPQGAAPQPHPPAFSAFNLYWDDPPERGAPSTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255
```

RESULT 13

AAM51143
ID AAM51143 standard; Protein; 1255 AA.

XX AC AAM51143;

XX DT 17-JUN-2002 (first entry)

XX DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX

KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KM tyrosine kinase; receptor; c-erbB2; gene therapy.

OS Homo sapiens.

PH Key Location/Qualifiers

FT Domain 1..653

FT /note= "extracellular domain"

FT Domain 676..1255

FT /note= "intracellular domain"

FT Domain 990..1255

FT /note= "phosphorylation domain"

PN WO200212341-A2.

PD 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT or enhancing an immune response to the protein, has Her-2/neu
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain -
XX

PS Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines
CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltaPD fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.0%; Score 6702; DB 23; Length 1255;

Best Local Similarity 98.3%; Pred. No. 0;

Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;

Qy 1 MELAALCRWGLLALLPFGAASQVCTGTDMKRLPASPTHELDMLRHLVQGCQVQGNL 60

Db 1 MELAALCRWGLLALLPFGAASQVCTGTDMKRLPASPTHELDMLRHLVQGCQVQGNL 60

```
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLREQLRSL TELKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
DB 121 DPLNNTTPVTGASPGGLREQLRSL TELKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
QY 181 LTLIDTNRSRACHPCSPMKGRSGCWGSESSDCQSLTRTVCAAGCARGCKPLPTDCCHEQC 240
DB 181 LTLIDTNRSRACHPCSPMKGRSGCWGSESSDCQSLTRTVCAAGCARGCKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNFTVSWLRVPKVSASHLEYTFGAS 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNFTVSWLRVPKVSASHLEYTFGAS 300
QY 301 CVTACPNYLSTDVGSCTLCPLHNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVR 360
DB 295 CVTACPNYLSTDVGSCTLCPLHNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVR 354
QY 361 AVTSANTQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTELBEITGYLYISA 420
DB 355 AVTSANTQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTELBEITGYLYISA 414
QY 421 WPSLPLDSVFQNLQVIRGRILHNGAYSLTLOGLISWLGRLSRLGSLALIHNTHL 480
DB 415 WPSLPLDSVFQNLQVIRGRILHNGAYSLTLOGLISWLGRLSRLGSLALIHNTHL 474
QY 481 CFVHTVPWDOLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQF 540
DB 475 CFVHTVPWDOLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQF 534
QY 541 LRGOECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 600
DB 535 LRGOECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 594
QY 601 FCVAPCSGKVPDLSYMPIWKFPDEBEGACQCPINCTHSCVDLDDKCPAEQASPLTISI 660
DB 595 FCVAPCSGKVPDLSYMPIWKFPDEBEGACQCPINCTHSCVDLDDKCPAEQASPLTISI 654
QY 661 VSAVVGILLVVLGVVFGIILKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQOMRI 720
DB 655 ISAVVGILLVVLGVVFGIILKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQOMRI 714
QY 721 LKETELRKVKVLSGAGFTYVKGIIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVM 780
DB 715 LKETELRKVKVLSGAGFTYVKGIIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVM 774
QY 781 AGVGSPPVSRLLGICLTSTVQLVTLMPYCCLLDHVRENRGRGLSQDOLLNWCQIAKMS 840
DB 775 AGVGSPPVSRLLGICLTSTVQLVTLMPYCCLLDHVRENRGRGLSQDOLLNWCQIAKMS 834
QY 841 YLEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGSKVPKIKWMALESI 900
DB 835 YLEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGSKVPKIKWMALESI 894
QY 901 LRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 960
DB 895 LRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 954
QY 961 MVKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDEM 1020
DB 955 MVKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDEM 1014
QY 1021 GDLVDAEEVLVPOQGFCDPAPAGWVHRRSSSTRSGGDLTLTGLEPSEEEAPRSP 1080
DB 1015 GDLVDAEEVLVPOQGFCDPAPAGWVHRRSSSTRSGGDLTLTGLEPSEEEAPRSP 1074
QY 1081 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSFLQRYSEDPTVPLPSETDGVVAPLTCS 1140
DB 1075 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSFLQRYSEDPTVPLPSETDGVVAPLTCS 1134
QY 1141 POPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAGGAVENP 1200
```

```
DB 1135 POPEYVNPQDVRPQPPSPREGPLPAARPAGATLERPKTILSPGKNGVVKDVFAGGAVENP 1194
QY 1201 EYLTPQCGAAPQHPHPPAFSPAFDNLVYWOODPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
DB 1195 EYLTPQCGAAPQHPHPPAFSPAFDNLVYWOODPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
QY 1261 V 1261
DB 1255 V 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
DR WPI; 2002-280741/32.
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide.
PS Disclosure; Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.0%; Score 6702; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;

QY 1 MELAALCRGGLLALLLPAGAASTQVCTGTDMLKRLPASPEHLDMLRLHYQGQVVOGNNL 60
DB 1 MELAALCRGGLLALLLPAGAASTQVCTGTDMLKRLPASPEHLDMLRLHYQGQVVOGNNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
```

```
Db 61 ELTYLPTNASLSFLQDIOEVGYVLI AHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVTCASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGSESEDCQSILTRTVACAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGSESEDCQSILTRTVACAGGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPRKSDCLACLFHNSGICELHCPALVFNNFTVSFWLRVPKVSASHLERYTFGAS 300
Db 241 AAGCTGPRKSDCLACLFHNSGICELHCPALVYNTYNTDTFESMPNPE-----GRYTFGAS 294
Qy 301 CVTACPNYVLTSDVGSCTVCLPLHNOEVTADGTORCEKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYVLTSDVGSCTVCLPLHNOEVTADGTORCEKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANIOEFAGCKKI FGSFLAFLESFDDPASNTAPLOPELOQVFEITLBEITGYLYTISA 420
Db 355 AVTSANIOEFAGCKKI FGSFLAFLESFDDPASNTAPLOPELOQVFEITLBEITGYLYTISA 414
Qy 421 WPDLSPLDSVFQNLQVIRGRILHNGAYSILTLOGLGISWGLRSLRSLRSGLALIHNNTHL 480
Db 415 WPDLSPLDSVFQNLQVIRGRILHNGAYSILTLOGLGISWGLRSLRSLRSGLALIHNNTHL 474
Qy 481 CFVHTVPHDOLFRNPHOALLHTANRPEDECVGEGGLACHQLCARHCWGPGTQCVCNCSQF 540
Db 475 CFVHTVPHDOLFRNPHOALLHTANRPEDECVGEGGLACHQLCARHCWGPGTQCVCNCSQF 534
Qy 541 LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 600
Db 535 LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 594
Qy 601 FCVARCPGKVPDLSYMPIMKFPDDEGACQPCINCTHSCVDLDDKGCPEAQRASPLTSI 660
Db 595 FCVARCPGKVPDLSYMPIMKFPDDEGACQPCINCTHSCVDLDDKGCPEAQRASPLTSI 654
Qy 661 VSAVGILLVVVLGVVVGILIKRROOKIRKTYMRELLQETELVEPLTPSGAMPNQAQMRI 720
Db 655 ISAVVGILLVVVLGVVVGILIKRROOKIRKTYMRELLQETELVEPLTPSGAMPNQAQMRI 714
Qy 721 LKETELRKVKVGLSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 780
Db 715 LKETELRKVKVGLSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 774
Qy 781 AGVGSPPYVSRLLGICLTSTVQLVTQMPYGCILLDHVRENRRGLSGQDLLNWCMIQAKGMS 840
Db 775 AGVGSPPYVSRLLGICLTSTVQLVTQMPYGCILLDHVRENRRGLSGQDLLNWCMIQAKGMS 834
Qy 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYHADGKVPKKNWALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYHADGKVPKKNWALESI 894
Qy 901 LRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPOPPICITIDVYMI 960
Db 895 LRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPOPPICITIDVYMI 954
Qy 961 MVKCMWIDSECRPRELVSFSPRMARDPQRFVVIQNEDLGPASPLDSTFYRSLEDDDM 1020
Db 955 MVKCMWIDSECRPRELVSFSPRMARDPQRFVVIQNEDLGPASPLDSTFYRSLEDDDM 1014
Qy 1021 GDLDVAEYLVPPQGGFFCPDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db 1015 GDLDVAEYLVPPQGGFFCPDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGMAAGKLSLPTHDPSPLQRYSEDPVLPSETDGYVAPLTCSS 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAGKLSLPTHDPSPLQRYSEDPVLPSETDGYVAPLTCSS 1134
Qy 1141 PQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVWKDVFAFGAVENP 1200
```

```
Db 1135 PQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERPKTSLPGKNGVWKDVFAFGAVENP 1194
Qy 1201 EYLTPQGGAAAPQPPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPQGGAAAPQPPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN W09316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU ) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
DR MPI, 1993-272889/34.
DR N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PS antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells. such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 97.4%; Score 6659; DB 14; Length 1433;
Best Local Similarity 97.8%; Pred No. 0;
Matches 1233; Conservative 5; Mismatches 17; Indels 6; Gaps 1;

Qy 1 MELAAALCRWGLLLALLPFGAASSTQVCTGCTDMKRLRLPASPETHLMLRHLHYQCVVQGNL 60
Db 1 MELAAALCRWGLLLALLPFGAASSTQVCTGCTDMKRLRLPASPETHLMLRHLHYQCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVGYVLI AHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVGYVLI AHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVTCASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGSESEDCQSILTRTVACAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGSESEDCQSILTRTVACAGGCARCKGPLPTDCCHEQC 240
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds
(without alignments)

5267.077 Million cell updates/sec

Title: SEQ4-325-339-14

Perfect score: 6852

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 6730 | 98.2 | 1255 | 1 A24571 | protein-tyrosine k |
| 2 | 5925 | 86.3 | 1260 | 1 TVRTNU | protein-tyrosine k |
| 3 | 5916.5 | 86.3 | 1254 | 2 I48161 | p-185 precursor |
| 4 | 3150 | 46.0 | 1210 | 1 GQHUE | epidermal growth f |
| 5 | 3123 | 45.6 | 1210 | 2 A53183 | epidermal growth f |
| 6 | 3099.5 | 45.2 | 1223 | 1 TVCHLV | epidermal growth f |
| 7 | 2981.5 | 43.5 | 1308 | 2 A47233 | epidermal growth f |
| 8 | 2672 | 39.0 | 1166 | 1 S06142 | protein-tyrosine k |
| 9 | 2430.5 | 35.5 | 1342 | 2 A36223 | kinase-related tra |
| 10 | 2342.5 | 34.2 | 1339 | 2 JC4387 | epidermal growth f |
| 11 | 1766.5 | 25.8 | 698 | 1 TVFVLV | protein-tyrosine k |
| 12 | 1703 | 24.9 | 604 | 1 TVYUHV | protein-tyrosine k |
| 13 | 1647 | 24.0 | 544 | 2 S35745 | protein-tyrosine k |
| 14 | 1643.5 | 24.0 | 1330 | 1 QQFFE | epidermal growth f |
| 15 | 1640 | 23.9 | 545 | 2 S00727 | kinase-related tra |
| 16 | 1623 | 23.7 | 540 | 2 B44776 | protein-tyrosine k |
| 17 | 1621 | 23.7 | 540 | 1 TVFVBB | protein-tyrosine k |
| 18 | 1515 | 22.1 | 644 | 2 A36325 | epidermal growth f |
| 19 | 1300 | 19.0 | 1323 | 2 E88257 | protein let-23 lim |
| 20 | 1300 | 19.0 | 1374 | 2 S70712 | protein-tyrosine k |
| 21 | 1214 | 17.7 | 1369 | 2 S70713 | protein-tyrosine k |
| 22 | 1180 | 17.2 | 1717 | 1 A45558 | epidermal growth f |
| 23 | 1131 | 16.5 | 527 | 2 A42032 | epidermal growth f |
| 24 | 988.5 | 14.4 | 843 | 2 A27131 | epidermal growth f |
| 25 | 806.5 | 11.8 | 346 | 2 S13807 | protein-tyrosine k |
| 26 | 754.5 | 11.0 | 311 | 2 S13808 | protein-tyrosine k |
| 27 | 730 | 10.7 | 1363 | 2 T43220 | insulin-like growt |
| 28 | 698 | 10.2 | 1382 | 1 INHUR | insulin receptor p |
| 29 | 693 | 10.1 | 1607 | 2 T43212 | insulin-like growt |

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N:Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erb

C:Species: Homo sapiens (man)

C>Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C:Accession: A24571; A25491; A44188; B4188; I59509; I57622

R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth

A:Reference number: A24571; MUID:86118663; PMID:3003577

A:Accession: A24571

A:Molecule type: mRNA

A:Residues: 1-1255 <YAM>

A:Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epider

A:Reference number: A25491; MUID:86016729; PMID:2995967

A:Accession: A25491

A:Molecule type: DNA

A:Residues: 737-1031 <SEM>

A:Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R:Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,

Science 230, 1132-1139, 1985

A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromo

A:Reference number: A44188; MUID:86070181; PMID:2999974

A:Accession: A44188

A:Molecule type: DNA

A:Residues: 740-910 <COU1>

A:Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A:Accession: B44188

A:Molecule type: mRNA

A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A:Cross-references: GB:M11730; NID:G183986

R:King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A:Reference number: I59509; MUID:85272597; PMID:2992089

A:Accession: I59509

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 832-909 <REX>

A:Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptiona

A:Reference number: I57622; MUID:87286898; PMID:3039351

A:Accession: I57622

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGU: HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F:22-653/Domain: extracellular #status predicted <EXT>
F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
F:68.124.187.259.530.571.629/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.2%; Score 6730; DB 1; Length 1255;
Best Local Similarity 98.4%; Pred. No. 1.6e-267;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASTVCTGDMKRLPASPETHLDMRLHLYGCGVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASTVCTGDMKRLPASPETHLDMRLHLYGCGVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVGYVLIHANOVRQVPLQRLIRVGTQLPEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVGYVLIHANOVRQVPLQRLIRVGTQLPEDNYALAVLDNG 120

Qy 121 DPLNNTTPVTGASPGGLRELQRLSITLKGVLIRQNPOLCYQDTILWKDIFHNKOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSITLKGVLIRQNPOLCYQDTILWKDIFHNKOLA 180

Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCHEQC 240

Qy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPLNHQNQVTAEDGTQRCCKSKPCARVCYGLGFNNFTVSFWLR-V 359
Db 301 YNYLSTDVGSCTLVCPLNHQNQVTAEDGTQRCCKSKPCARVCYGLGFNNH-----LREV 353

Qy 360 PKVSASHLEFPAGCKKIFGSLAFIPESFDGDPASNTAPLQEQIQVFETLEEITGYLYIS 419
Db 354 RAVTSANIQEPAGCKKIFGSLAFIPESFDGDPASNTAPLQEQIQVFETLEEITGYLYIS 413

Qy 420 AWPDSLPLSVFQNLQVIRGRILHNGAYSITLQGLGISWGLRLSLRELGLALIHNTHT 479
Db 414 AWPDSLPLSVFQNLQVIRGRILHNGAYSITLQGLGISWGLRLSLRELGLALIHNTHT 473

Qy 480 LCFVHTVPWQDLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTQCVNCSQ 539
Db 474 LCFVHTVPWQDLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTQCVNCSQ 533

Qy 540 FLRQECVEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCAVACHYKDP 599
Db 534 FLRQECVEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCAVACHYKDP 593

Qy 600 PFCVARCPSPGKPDLSYMPITWKFDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTS 659
Db 600 PFCVARCPSPGKPDLSYMPITWKFDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTS 659

Db 594 PFCVARCPSPGKPDLSYMPITWKFDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTS 653
Qy 660 IVSAVVGTLAAVVLGVVFGILLIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQHR 719
Db 654 IISAVVGTLAAVVLGVVFGILLIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQHR 713
Qy 720 ILKETELRKVKVLGSGAGFTYVKGWIPDGENVAPIVAIKVLRRENTSPKANKEILDEAYV 779
Db 714 ILKETELRKVKVLGSGAGFTYVKGWIPDGENVAPIVAIKVLRRENTSPKANKEILDEAYV 773
Qy 780 MAGVSPYVSRLLGICLTSTVQLVTLQMPYCCLLDHVRENRRGLSGDQLLNWCQIAKGM 839
Db 774 MAGVSPYVSRLLGICLTSTVQLVTLQMPYCCLLDHVRENRRGLSGDQLLNWCQIAKGM 833
Qy 840 SYLEDVRLVHRDLAARNVLKSPNHVKTDFGLARLLDIDETEVHADCGKVPKRWMALES 899
Db 834 SYLEDVRLVHRDLAARNVLKSPNHVKTDFGLARLLDIDETEVHADCGKVPKRWMALES 893
Qy 900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYM 959
Db 894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYM 953
Qy 960 IMVKCWMIDSECRPRPRELVESEFMRMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDD 1019
Db 954 IMVKCWMIDSECRPRPRELVESEFMRMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDD 1013
Qy 1020 MGDLDVAEEYLVPOOGFFCDDPAPGAGMWHRRSSSTRSGGDLTLGLPSEEEAPRS 1079
Db 1014 MGDLDVAEEYLVPOOGFFCDDPAPGAGMWHRRSSSTRSGGDLTLGLPSEEEAPRS 1073
Qy 1080 PLAPSEGAGSDVFDGLGMGAAGLQSLPHTDPSPLORYSDPTVPLPSETDGYVAPLTC 1139
Db 1074 PLAPSEGAGSDVFDGLGMGAAGLQSLPHTDPSPLORYSDPTVPLPSETDGYVAPLTC 1133
Qy 1140 SPOEYVNDVRPQPPSPREGPLPAARPAAGATLAKTLSPGKXGVKVDVAFGAGAVEN 1199
Db 1134 SPOEYVNDVRPQPPSPREGPLPAARPAAGATLAKTLSPGKXGVKVDVAFGAGAVEN 1193
Qy 1200 PEYLTPOGGAAPHPHPPAPFPAFNDLYYMDQDPPERGAPBPTFKGTPTAENPEYLGIDV 1259
Db 1194 PEYLTPOGGAAPHPHPPAPFPAFNDLYYMDQDPPERGAPBPTFKGTPTAENPEYLGIDV 1253
Qy 1260 PV 1261
Db 1254 PV 1255

RESULT 2
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) new precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The new oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat new oncogene transmembrane domain reveals no
2-thiaxolylformamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663, 'v', 665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F;658-680/Domain: transmembrane #status predicted <TMN>
F;723-988/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;71,191,263,535,576,634/Binding site: carbonylrate (Aen) (covalent) #status predicted
F;691/Binding site: phosphate (Thr) (covalent) #status predicted
F;758/Active site: Lys #status predicted
F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.5%; Score 5925; DB 1; Length 1260;
Best Local Similarity 86.6%; Pred. No. 1.1e-234;
Matches 1095; Conservative 56; Mismatches 103; Indels 10; Gaps 4;

Qy 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHLYGCGVQVQNL 60
Db 4 MELAAMCRWGLLALLPPGIAGTQCTGCTDMKRLPASPETHLDMRLHLYGCGVQVQNL 63

Qy 61 ELTYLPTNASLSFLQDIOEVGYVLIHANNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 64 ELTYVPTNASLSFLQDIOEVGYVLIHANNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNR 123

Qy 121 DPLNNTTPVT-GASPGGLRELQSLTEILKGGVLIQNPOLCYQDILWKDIFHKNQL 179
Db 124 DPQNVAASTPGRTPGELRELQSLTEILKGGVLIQNPOLCYQDILWKDIFHKNQL 183

Qy 180 ALTLDITNRSRACHPCSPMKGSGESSDCSLTRTVCCAGCARCKGLPLTDCCHEQ 239
Db 184 APVDITNRSRACHPCPCAPACKDNHCGESPDCQILGTICTSCACRCKGLPLTDCCHEQ 243

Qy 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTAC 299
Db 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTTC 303

Qy 300 PYNVLSIDVGSCTLVCPHMQEVAEDGTORCEKSPCARVCYGLGMFNFTVSWFLRV 359
Db 304 PYNVLSIDVGSCTLVCPHMQEVAEDGTORCEKSPCARVCYGLGMFNFTVSWFLRV 359

Qy 360 PK-VSASHLEEFACCKIFGSLAFPLPSFDGDPASNTAPLOPEOLQVFTLEETGYLYI 418
Db 357 ARAITSDNVQEFDCCKIFGSLAFPLPSFDGDPASNTAPLOPEOLQVFTLEETGYLYI 416

Qy 419 SAMPDLSLVSFONQVIRILHNGAYSLTQGLGISMGLRSLRSLGSLALIHNT 478
Db 417 SAMPDLSLVSFONQVIRILHNGAYSLTQGLGISMGLRSLRSLGSLALIHNA 476

Qy 479 HLCFVHTVPMDQFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGCHGPGTQVCNC 537
Db 477 HLCFVHTVPMDQFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGCHGPGTQVCNC 536

Qy 538 SFLRGQECVEECRVLQGLPREYVNAHRLCPCHPECPQNGSVTCFGEADQCVACAHYK 597
Db 537 SHFLRGQECVEECRVLQGLPREYVNAHRLCPCHPECPQNGSVTCFGEADQCVACAHYK 596

Qy 598 DPFCVACRCSGKVPDLSYMPIWKFPDEGACQPCINCTHSCVDLDDKGCAPQASPL 657
Db 597 DSSSCVACRCSGKVPDLSYMPIWKFPDEGACQPCINCTHSCVDLDDKGCAPQASPL 656

Qy 658 TSIVSAVVGILLVVLGWFGLIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQ 717
Db 657 TFIATVEGVLILLVVLGWFGLIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQ 716

Qy 718 MRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVRENTSPKANETILDEA 777
Db 717 MRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVRENTSPKANETILDEA 776

Qy 778 YVMAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGRGSDLLNWCQIAK 837
Db 777 YVMAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGRGSDLLNWCQIAK 836

Qy 838 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMAL 897

Db 837 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMAL 896
Qy 898 ESIILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDV 957
Db 897 ESIILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDV 956

Qy 958 YMIWVKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLED 1017
Db 957 YMIWVKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLED 1016

Qy 1018 DDMGDLVDAEYLVPQGFPCPDPAAGAGMWHHRHSSSTRSGGDLTLGLPESEBAP 1077
Db 1017 DDMGDLVDAEYLVPQGFPCPDPAAGAGMWHHRHSSSTRSGGDLTLGLPESEBAP 1076

Qy 1078 RSLPAPSEGAGSDVFDGDLGMAAKGLQSLTPHDSPLQRYSEDPVLPSETDGYVAPL 1137
Db 1077 RSLPAPSEGAGSDVFDGDLGMAAKGLQSLTPHDSPLQRYSEDPVLPSETDGYVAPL 1136

Qy 1138 TCSPOPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVWKDVFAFGGAN 1197
Db 1137 ACSPOPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVWKDVFAFGGAN 1196

Qy 1198 ENPEYLTPOGGAAPQHPHPPAFSPAFDNLYTWDQDPPERGAAPPSTFKGTPTAENPEYI 1257
Db 1197 ENPEYLTPOGGAAPQHPHPPAFSPAFDNLYTWDQDPPERGAAPPSTFKGTPTAENPEYI 1256

Qy 1258 DVPV 1261
Db 1257 DVPV 1260

RESULT 3
148161
P-185 precursor - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C;Accession: 148161
R;Nakamura, T.; Uehijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa
Gene 140, 251-255, 1994
A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A;Reference number: 148161; MUID:94193007; PMID:7908275
A;Accession: 148161
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1254 <RES>
A;Cross-references: GB:D16295; NID:G493236; PIDN:BAA03801.1; PID:G747595
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif

Query Match 86.3%; Score 5916.5; DB 2; Length 1254;
Best Local Similarity 86.3%; Pred. No. 2.4e-234;
Matches 1089; Conservative 66; Mismatches 98; Indels 9; Gaps 3;

Qy 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHLYGCGVQVQNL 60
Db 1 MELAAMCRWGLLALLPPGIAGTQCTGCTDMKRLPASPETHLDMRLHLYGCGVQVQNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVGYVLIHANNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVGYVLIHANNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNR 120

Qy 121 DPLNNTTPVTGASPGGLRELQSLTEILKGGVLIQNPOLCYQDILWKDIFHKNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQSLTEILKGGVLIQNPOLCYQDILWKDIFHKNQLA 180

Qy 181 LTLIDNRSRACHPCSPMKGSGESSDCSLTRTVCCAGCARCKGLPLTDCCHEQ 240
Db 181 PVDITNRSRACHPCPCAPACKDNHCGESPDCQILGTICTSCACRCKGLPLTDCCHEQ 240

R; Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A; Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
 A; Reference number: A60143; MUID: 85182650; PMID: 2985580
 A; Accession: A60143
 A; Molecule type: protein
 A; Residues: 740-744, 'X', 746-747 <RUS>
 R; Mroczkowski, B.; Mosig, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A; Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
 A; Reference number: A38023; MUID: 84191554; PMID: 6325948
 A; Contents: annotation; receptor activity
 A; Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R; Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A; Title: Functional independence of the epidermal growth factor receptor from a domain 2
 A; Reference number: A33331; MUID: 90003233; PMID: 2790960
 A; Contents: annotation; internalization signal
 C; Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor
 C; Genetics:
 A; Gene: GDB: EGFR
 A; Cross-references: GDB: 120610; OMIM: 131550
 A; Map position: 7p12.3-7p12.1
 C; Superfamily: epidermal growth factor receptor; protein kinase homology
 C; Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F: 1-24/Domain: signal sequence #status predicted <SIG>
 F: 25-645/Domain: extracellular #status predicted <EXT>
 F: 75-300/Domain: EGF receptor extracellular domain repeat <BE1>
 F: 390-600/Domain: EGF receptor extracellular domain repeat <BE2>
 F: 646-668/Domain: transmembrane #status predicted <TM>
 F: 669-1210/Domain: intracellular #status predicted <INT>
 F: 710-975/Domain: protein kinase homology <KIN>
 F: 718-726/Region: protein kinase ATP-binding motif
 F: 999-1046/Region: coated-pit mediated internalization signal
 F: 1047-1210/Region: inhibitory
 F: 128, 175, 352, 413, 444, 528, 603/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F: 745/Active site: Lys #status experimental

Query Match 46.0%; Score 3150; DB 1; Length 1210;
 Best Local Similarity 49.6%; Pred. No. 1.7e-121;
 Matches 631; Conservative 178; Mismatches 32; Indels 112; Gaps 23;

| | | | |
|----|-----|--|-----|
| Qy | 11 | LLALLPPGAA--STOVCTGDMKRLPASPETHDMLRLHYQCCVQVQGNLEITYLPTN | 68 |
| Db | 14 | LLAALCPASRALEBKVCQGTNSKLTQGTGFEDHFLQRMFNCEVVLGNLEITYYQNR | 73 |
| Qy | 69 | ASLSFLQDIOEVQGYVLIHQNVRQVPLQRLIRVGTQQLFEDNVALAVLQNGDPLNNTTP | 128 |
| Db | 74 | YDLSFLKTIQEVAGYVLIALTVERIPLENLQIIRGNMYENSVALAVLSND----- | 126 |
| Qy | 129 | VTGASPGGLRELQRLSRLTEILKGGVLIQRPQLCYQDTILWKOIFHKNNQLALTLIDTNR | 188 |
| Db | 127 | ---ANKTGLKELPMRNQIQLHGAVFPSNNPACNCVSIQWRDITVSSDFLSNMSDFQNH | 183 |
| Qy | 189 | SRACHPCSPMKSGRCGSESSDCQSLTRTVACGGCA-RCKGPLPTDCHEQCAAGCTGP | 247 |
| Db | 184 | LGSCQKCDPCSPNGSCMGAEENCCQKTKIQAQCCSGRCGRKSPSCCHNQCAAGCTGP | 243 |
| Qy | 248 | KHSDCLACLFHNSGICELCHPCALVTYNTDTFFSMNPGRYTFGASCVTACPNYLSLD | 307 |
| Db | 244 | RESDCLVCRKFRDEATCKDCTPPLMLNPTTYQMDVNPCKYSGFATCVKCKPRNYVTD | 303 |
| Qy | 308 | VGSCTLVCLPHNQEVTAEDGTQRCCKSKFCARVCYGLGNFNFTVFWLRVPRKVSASHL | 367 |
| Db | 304 | HGSCVRACGADSYEM-EEDGVRRCKCKCEGPCRCVNGIGI-GBFKDSL-----SINATNI | 356 |
| Qy | 368 | EEFAGCKKIFGSLAFIPESFDGPPASNTAPLQPEQLQVFTLEITCYLYISAWPSLDP | 427 |
| Db | 357 | KHFVNTSISGDHLIPVAFRGDSFTHPPDDQELDKVTKVEITGFLLIQAWPENRTD | 416 |
| Qy | 428 | LSVFONLQVIRGIRLHNGAYSLLTQGLIGISWLGSLRLSRLGSLALHNNHLCFCVHTVP | 487 |

RESULT 5

A53183

epidermal growth factor receptor precursor - mouse

C; Species: Mus musculus (house mouse)

C; Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999

C; Accession: A53183; A43818; S24942; A28941; S45325; I49643

R; Luettke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.; Genes Dev. 8, 399-413, 1994

A; Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor

A; Reference number: A53183; MUID: 94170986; PMID: 8125255

A; Accession: A53183

A; Molecule type: mRNA

A; Residues: 1-1210 <LUE>

A; Cross-references: GB: U03425

RESULT 6
 TUCHLV
 epidermal growth factor receptor precursor - chicken
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
 C:Species: Gallus gallus (chicken)
 C:Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
 C:Accession: A27720; A00643
 R:Laux, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennart
 Mol. Cell. Biol. 8, 1970-1978, 1988
 A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
 A:Reference number: A27720; MUID:88261272; PMID:3260329
 A:Accession: A27720
 A:Molecule type: mRNA
 A:Residues: 1-1223 <LAX>
 A:Cross-references: GB:M20386
 R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
 Cell 41, 719-726, 1985
 A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
 A:Reference number: A00643; MUID:85228222; PMID:2988784
 A:Accession: A00643
 A:Molecule type: mRNA
 A:Residues: 585-1223 <NIL>
 A:Cross-references: GB:M10066
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
 specific protein kinase
 F:1-30/Domain: signal sequence #status predicted <SIG>
 F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
 F:31-654/Domain: extracellular #status predicted <EXT>
 F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
 F:597-610/Domain: EGF receptor extracellular domain repeat <EE2>
 F:655-677/Domain: transmembrane #status predicted <TM>
 F:678-1223/Domain: intracellular #status predicted <INT>
 F:719-984/Domain: protein kinase homology <KIN>
 F:727-735/Region: protein kinase ATP-binding motif
 F:136-202/280: 361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
 F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:754/Active site: Lys #status predicted
 F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.2%; Score 3099.5; DB 1; Length 1223;
 Best Local Similarity 48.3%; Pred. No. 1.9e-119;
 Matches 630; Conservative 177; Mismatches 345; Indels 151; Gaps 26;

QY 8 RMGLLALLPPGAA-----STQVCTGTDMLRLPASPETHLMDLRHLVQGVQGVNLE 61
 DB 13 RGAALVLLGLVALCSAVEKKVCGTNNKLTQLGHVEDHFTSLQRMYNNECVLSLE 72
 QY 62 LTYLPTNASLSFLODIQEVGYVLIAHNOVQVQLRIRVGTQLFEDNYALAVLDNGD 121
 DB 73 ITYVHNRLDTLTKTIQEVAGVYVLIANMVDVIPLENLQIRGNVLYDNSFALVLSNVH 132
 QY 122 PLANVTPTVGASPGIGRELQRLSLTEILKGGVLIORNPOLCVQDTILMKDIFPHKNQAL 181
 DB 133 -MNKTQ-----GLRELPMSEILNGVYKISNNPKLNMDTVLMNDIIDTSRK-PL 182
 QY 182 TLID-TNRSRACHPCSPMKSGSRGSESSDCQSLTRTVCAGGCA-RCKGPLPTCCHEQ 239
 DB 183 TVLDFASNLSSCPKCHPCTEDHCHGAGEQNCQTILTKVICAQCSGRCKRGVPSCCCHNQ 242
 QY 240 CAAGCTGPKHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
 DB 243 CAAGCTGPGRESCLACRFRDQATCKDTCPLVLYNPITYQMDVNPPEGKYSFGATCVREC 302
 QY 300 PYNLYSTDVGSCTVCPHLNQHVTAEDGTQRECKSKPCARVCYGLGMFNPNFTVFWLRV 359
 DB 303 PHNVVYVTDHGSCVRSNCTDTVEV-EENGVRKCKKCDGLCKVCNGIG-----IGELGI 355
 QY 360 PKVSASHLEEFAGCKIFFGSLAFLPESPDGDPASNTAPLQEPQLQVFTLEETLYLIS 419

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epiderma.

A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <PLOS>
A;Cross-references: GB:U07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A;Note: sequence extracted from NCBI backbone (NCBIP:126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

Query Match 43.5%; Score 2981.5; DB 2; Length 1308;
Best Local Similarity 45.2%; Pred. No. 1.3e-114;
Matches 611; Conservative 187; Mismatches 375; Indels 179; Gaps 29;

QY 9 WGLLLALLPGAA-----STOVCTGDMKRLPASPETHLMDLRLHLYGCGVVOGNLELT 64
DB 8 WVVSVLLVAAGTVQPSDSQSCVAGTENKLSLSLDEQQYRALRYKRYENCEVVMGNLEITS 67
QY 65 LPTNASISFLQDIOEVGYVLIHNOVQVPLQRLIRVGTOLPEQNYALAVLDNGDPLN 124
DB 68 IEHRDLSPFURSREVTGYVVALNOFYLPLENRLIRGTUKLYEDRYALAIFLNVRKDG 127
QY 125 NTPPTVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQALTLI 184
DB 128 NF-----GLQELGKLNLTILNGVYVDQNKFLCYADTIHWQDIVRNPPSNLTIV 178
QY 185 DTVRSRACHPCSPMKGSRGWSSEDCQSLTRTVACAGC-ARCKGPLPTDCHEQCAAG 243
DB 179 STNGSSGCRCHKSCCTG-RCWGPTEHNHCOTLTRTVCAEQDGRGYPVSDCHRECAAG 237
QY 244 CTGPKHSDCLACHENHSGICELHCPALVTYNTDTESMPNPEGRYTFGASCVTAPYNY 303
DB 238 CSGPKOTDCFACWNFDSGACVTCQPTQTFYNTPTTFOLEHFNKAYTYGAFCKKCPHF 297
QY 304 LSTDVGSCTLVCLPHNQVTAEDGTQRCCKSPCARVCYGLGMFNNFTVSFLRVPKVS 363
DB 298 V-VDSSSVCRACFPSSKMEV-EENGKMKPCCTDICKKACDGI-----TGLMSAQTV 349
QY 364 ASHLEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPELOVETLEETIGYLYISAMPD 423
DB 350 SSNIDPFNCTKINGNLFIPLVTGIGHGDPYNAIRAIDPEKLVNFRVREITGFLNIQSWPP 409
QY 424 SLPLSVFONQLOVIRGRILHNGAYSLTLOGLGTSWLGRLSLRELGLGSLALIHNTLCFV 483
DB 410 NMTDFSVFSLNVTIGRVLVSGLSLLIKQGLTSLOFSLKISAGNIYITDNSNLCY 469
QY 484 HTVPMDOLFRPNHQAALLHTANRPEDECVGEGLAACHQLCARGHCGPGPTQCVNCSQFLRG 543
DB 470 HTINWTLFTSTINQIRIVIRDNRAENCTAEGVMCNHLCSDDGCGWPGPDQCLSCRRFSR 529
QY 544 QECVECRVLQGLPREVYNARHCLPCHPEQCP-ONGSVTCFGEADOCVACAHYKDPFFC 602
DB 530 RICIESCNLYDGFREFENGISICEVDQCEKMEDEGLLTCHGFGPDNCTKCSHFQDGPNC 589
QY 603 VAPCPSGVKPDLSYMPITWKEPDEEGACOPCPINCTHSCVDLDDKGC-----PAE 651
DB 590 VEKCPDGLQANF--IFKTDPRCHPCPNCTOCNGPTSHDCIYYPWTGHSTLPQH 647
QY 652 QRASPLTSIVSAVV-GILLYVVLGVVFGILIKRRQOKIRKYTMRLLIQETELVEPLTPSG 710
DB 648 AR-TPL--IAAGVIGLFIIVGLTFVAVVRRKSIK-KKRALRRFL-ETELVEPLTPSG 702
QY 711 AMPNQAMRLKTELAKVKVLGSGAGTGYKGIWIPDGENVKIPIVAKVLRNTPSKAN 770
DB 703 TAPNOAQLRLKETELKRVKVLGSGAGTGYKGIWVPEGETVPIVPAIKILNETTGPKAN 762
QY 771 KEILDEAYVAGVSPVSRLLGLTSTVOLVTOLMPYCLLDHVRNRRGLRGSQDLN 830
DB 763 VPMDEALINASHDHFLVRLGLVGLSPITQLVTQLMPHGLLEYVHEHKDNTGSQLLN 822
QY 831 WCMQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKV 890

DB 823 MCVOIAKGMYLSEERLVHRDLAARNVLKSPNHVKITDFGLARLLDSEGEKEYNADGKM 882
QY 891 PIKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQ 950
DB 883 PIKMALECIHYKFTHQSDVMSYGVTVWELMTFGGKPYDGIPTREIPDLLEKGERLPQ 942
QY 951 PICTIDVYIMVKCMWIDSECRPRFRELVSERFARMARDPQVFFVIONED-LGPASPLDST 1009
DB 943 PICTIDVYIMVKCMWIDADSBRKFKELAAAEFARMARDPQVLYVIOGDDRMKLPSPNDK 1002
QY 1010 FYRSLLEDDMDGLVDAEYLVPOQGFCCPDPAFGAGGMVHRRHSSTSRSGGDLTLGL 1069
DB 1003 FFQNLLEDEEDMDMAEEYLVLP-QAFNIPPP-----IYTSRARIIDNRS-----EIGH 1050
QY 1070 EPSEERAPRS-----PLAP-SEGAGSDVFDGDLGMA 1100
DB 1051 SPPPAYTPMSGNOFVYRDGGFAAEQGVSVPYRAPTSTIPEAPVAQAGATAEIFDDSCNGT 1110
QY 1101 AKGLQSLPTHDPSPLOQRYSEDPTVPLPS-----ETDGYVAPLTCSPQPEYVNPQDVRP 1153
DB 1111 LRKPVAHFVQEDSSTQRYSDPTVFAPERSPRGELDEEGYTMRMKPKOEYLYNPVE--- 1167
QY 1154 QPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTPOGCAAPQ 1213
DB 1168 -----ENPVSRR-----KXGDLQ-----ALDNEYHNASNG----- 1194
QY 1214 HPPPA-----FSPAFDNLYYWDQDPPPERGA- 1238
DB 1195 -PKAEDEVYNEPLNTFANTLGKAEYLNKNTLSMPEKAKAFDNDPDYWNHSLPPRSTL 1253
QY 1239 -PPSTFKGPT-----AENPEYL 1255
DB 1254 QHPDYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transform
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Zaulf, F.; Telling, A.; Rober
Nature 341, 415-421 1989
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu 1
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <WIT>
A;Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R;Adam, D.; Maeueller, W.; Scharf, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphoph
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C;Genetics:
A;Map position: Y
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; ty
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

Query Match 39.0%; Score 2672; DB 1; Length 1166;
Best Local Similarity 44.9%; Pred. No. 5e-102;

Db 580 RRGPHCVNCPHGILG--AKGPIYKYPDAQNECRPCHECTQCNCGFELQDCLGQAQVLM 637
QY 655 SPLTSTVSAVWGILLVVLGVWFGILIKRRQOKIR-KYTMRRLLQTELVEPLTPSGAMP 713
Db 638 SKPHLVIATVG--LAVILMILGSGFLYWRGRIQKRAMRRYLERGESIEPLDPS-EKA 694
QY 714 NQAMRILKETELRKVKVVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEI 773
Db 695 NKVLARIFKETELRKVLKVLGSGVFGTVHKGIWIPGESIKIPVICIVIEDKSGRQFOAV 754
QY 774 LDEAYMAGVGSPPVSRLLGICLTSTVOLTPMYPGCLLDHVHRENRGLSGODLNNCM 833
Db 755 THMLAVGSLDRAHIVRLGLCGSSQLQVTLPLGSLDHDVHQRETLGPOLLNMGV 814
QY 834 QIAKMSYLEDLVLRHDLAARNVLKSPNHVKITDGLARLLDIDETEYHADGGKVPVK 893
Db 815 QIAKMGWYLEEHSVHRDLARNVLMKSPSQVQVADFGVADLLPPDKQLLHSEAKTPK 874
QY 894 WHALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPIC 953
Db 875 WMALESIFGKYTHQSDVMSYGVTVWELMTFGAEPYAGURLABIPDLLEKGERLAQPOIC 934
QY 954 TIDVYMWKCMIDSECRPRELVSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRS 1013
Db 935 TIDVYMWKCMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GPGTP--PAAEPS 991
QY 1014 LLEDDDMGLVDAEEYLVPOQGFPCPDPAAGAGGVHHRHRSSTRSGGCDLTGLLEPSE 1073
Db 992 VLTKEL-----QEALEPEL-----DLDLLEAE 1017
QY 1074 E-----EAPRSLAPSEG-----AGSDVDFDGLGMAAKGLQ 1105
Db 1018 EGLATSLGALSALPTGLTRPGSQSLSPSSGYMPMNQSSLGEACLDASVLGREGQFSR 1077
QY 1106 SLPTHDPSPQLQRYSEDPTVLPSETDGVY----APL-----TC-----SPOPE---Y 1145
Db 1078 PISLH-P1PRGR-----PASESEGHWTGSEAELOEKVSVCRSRSRSPRPRGDSAY 1129
QY 1146 VNQPDVVRPOPSPREGP-----LPAARPAGATLERAKTLSP-GKNGYV----- 1187
Db 1130 HSQRHSLLPVTPSPPLGEEEDGNGYVMPDTHLRGASSSREGLTSSVGLSGVLTGEED 1189
QY 1188 KDVFAFGGAVENPEYLTPOGGAAPQHPHP 1216
Db 1190 ED-----EYEMNKRKRGSP-PRPP 1209
RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <NIL>
A:Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A:Note: in Genbank entry CHKERBBP, release 109.0, the source is designated as Gallus gal
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F:1-6/Product: gag protein (fragment) #status predicted <GAG>
F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F:194-459/Domain: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif

F:229/Active site: Lys #status predicted

Query Match 25.8%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 2.7e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;
QY 584 GPADOCVCAHYKDPFPCVAPCPGKPDLSYMPYWKFPDEGACQCPINCTHSCVDL 643
Db 60 GP--DRCMCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCQLCHPNCGRGKGP 116
QY 644 DDKGCPAEQASPLTSIVSAVW-GILLVVVLGVVGLIKRROOKIRKYTMRRLLQTEL 702
Db 117 GUEGCP--NGSKTPSIAAGVWGLCLLVVVGIGLYLRRR-HIVKRTLRLLRQLREL 172
QY 703 VEPLTPSGAMPNQAOMRILKETELRKVKVVLGSAFGTVYKGIWIPDGENVKIPVAIKVL 762
Db 173 VEPLTPSGAPNQAHLRIKETEFKVKVVLGSAFGTVYKGLWIPGEKVKIPVAIKELR 232
QY 763 ENTSPKANKEILDEAYMAGVGSPPVSRLLGICLTSTVOLTPMYPGCLLDHVHRENRGR 822
Db 233 EATSPKANKEILDEAYMASVDNPHVCRLLGICLTSTVOLITQMLPYGCLLDYIREHKDN 292
QY 823 LGSODLLNWCMIQAKMSYLEDLVLRDLAARNVLKSPNHVKITDGLARLLDIDETE 882
Db 293 IGSQYLLNMCVQIAKGMYLEERRLVHRLAARNVLKTPQHVKITDGLAKLLGADEKE 352
QY 883 YHADGGKVPKMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 942
Db 353 YHAEKGKVPKMALESILHRYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 412
QY 943 KGERLPQPICTIDVYMWKCMIDSECRPRELVSEFSRMARDPQRFVVIQ-NEDLG 1001
Db 413 KGERLPQPICTIDVYMWKCMIDADSRPKFRELIAEFSKWARDPPRYLVIOGDERMH 472
QY 1002 PASPLDSTFYRSLEDDMDGLVDAEEYLVPOQGFPCPDPAAGAGGVHHRHRSSTRSG 1061
Db 473 LPSPTDSKPYRTLMEBEDMEDIVDADEYLVPHQGF-----NSPST--- 513
QY 1062 CGDLTLGLSPSBEAERSPL-----APSEGAGSDVDFDGLGMAAKGLQSLPTHDPSPLO 1116
Db 514 -----SRTPLSLSATSNNSATNCID-----RNGQGHVPVREDSFVQ 550
QY 1117 RYSEDPVPLPSET--DGYVAPLTCSPQPEYVNPQDVRFPQPPSPREGPLPAARPAGATLE 1174
Db 551 RYSSDPTGNFLEESIDDGFL-----PAPEYVNYQ--LMPKKPS----- 585
QY 1175 RAKTSLSPGKGVVQKVF-----AFGGAIVENPEYLTPOGGAAPQHPHPAPF 1219
Db 586 ----TAMVQNIYNNISLTAISKLPMDSRYSQNSHSTAVDNPEYL-----NTNQSPLA 633
QY 1220 SPAFNNLYWDO-----DPE-----RGAPPSTFKGTPTAENPEYVLGLDVP 1260
Db 634 KTVFESSPYWIOSGNHQINLDNPDYQODFLPNETKPNGLLKVPAENPEYLRVAAP 689

RESULT 12

TVYUW

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C>Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fam
A:Reference number: A00644; MUID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R:Debuire, B.; Henry, C.; Benaissa, M.; Biserce, G.; Claverie, J.M.; Saule, S.; Martin
Science 224, 1456-1459, 1984
A:Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type o
A:Reference number: A38022; MUID:84223957; PMID:6328658

A:Accession: A38022
A:Molecule type: DNA
A:Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
A:Cross-references: GB:K02006
C:Genetics:

A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif
F:165/Active site: Lys #status predicted

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 9e-63;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

```
QY 593 CAHYKDPFCVACRCPGVKPDLSYMPIMKFPDEEGACQPCPINTCHSCVDLDDKCPAEQ 652
DB 3 CAHFDGPHCVKACPAVLGENDTL-VRKYADANAVCQLCHPNCNCTRGCKGPGLEGCP--- 58

QY 653 RASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGA 711
DB 59 NGSKTPSTAAGVVGGLLVVGLGILYLRER-HIVKRTLRLLQERELVEPLTPSGE 117

QY 712 MPNQMRILKETELRKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANK 771
DB 118 APQAHLRLKETEFKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANK 177

QY 772 EILDAYVMAGVSPVSRLLGICLTSTVOLTPQMPYGLLDHVRNRRGRSGOOLNM 831
DB 178 EILDAYVMASVDNPHVCRLLGICLTSTVOLTPQMPYGLLDYIREHKDNGSQVLLNM 237

QY 832 CMQIAKGSYLEDVLRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVP 891
DB 238 CVQIAKGMNLEERLVRDLAARNVVKTPQHVKITDFGLAKLLGADEKEYHAEGKVP 297

QY 892 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQP 951
DB 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQP 357

QY 952 ICTIDVYIMVKWCMIDSECRPFRELSEFSEMRARDPQFVVIO-NEDLGASPLDSTF 1010
DB 358 ICTIDVYIMVKWCMIDASRPKRELIAEFSEMRARDPQFVVIOQDERMHLPSPTDSKF 417

QY 1011 YRSLLEDDMDGLDAEYLVPOQGFCCPDPAFGAGMVHRRHSSTRSGGDLTLGLE 1070
DB 418 YRLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449

QY 1071 PSEEEAPRSP-----APSEGAGSDVFDGLGMGAAGKGLQSLPHTDPSPLQRYSEDPTVP 1125
DB 450 -----SRTPLLSLSATSNNATNCID-----RNCQGHFVREDSEFVQRYSSDPTGN 495

QY 1126 LPSET--DGYVAPLTCSPQPEYVQPDVVRPQPPSPREGPLPAARPAAGATLSPCK 1183
DB 496 FLEESIDDGFL-----PAPEVNO--LMPKPFSTAM----- 524

QY 1184 NGVVKDVFAP-----GGAVENPEYLTPOGGAAPQHPPPAPFAPD 1224
DB 525 --VQNIYNFISLITSLKPLMDSRYQNSHSTAVDNPYL-----NTNQSPKLTATVE 574

QY 1225 NLYYWDQDPPPERGAPPSTFKGTPTAENPEY 1254
DB 575 SSPYMTQSHNQ-----INLDNPDY 594
```

RESULT 13

S35745
Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Venustroem, B.
submitted to the EMBL Data Library, March 1993

A:Reference number: S35743

A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:

A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif
F:170/Active site: Lys #status predicted

Query Match 24.0%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.6e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

```
QY 584 GREADQCVACAHYKDPFCVACRCPGVKPDLSYMPIMKFPDEEGACQPCPINTCHSCVDL 643
DB 1 GP--DHCMKCAHFDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHPNCNCTRGCKGP 57

QY 644 DDKCPAQORASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMRRLLQETEL 702
DB 58 GLEGCP---NGSKTPSIAAGVVGGLLVVGLGILYLRER-HIVKRTLRLLQEREL 113

QY 703 VEPLTPSGAMPNQMRILKETELRKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLR 762
DB 114 VEPLTPSGEAPNQAHRLKETEFKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLR 173

QY 763 ENTPSKANKEILDEAYVMAGVSPVSRLLGICLTSTVOLTPQMPYGLLDHVRNRRGR 822
DB 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLTPQMPYGLLDYIREHKON 233

QY 823 LGSOLLNWCQIAKGSYLEDVLRDLAARNVVKSPNHVKITDFGLARLLDIDETE 882
DB 234 LGSQYLLNWCQIAKGMNLEERLVRDLAARNVVKTPQHVKITDFGLAKQLGADEKE 293

QY 883 YHADGGKVPKIKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLLE 942
DB 294 YHAEGKVPKIKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLE 353

QY 943 KGERLPQPPICTIDVYIMVKWCMIDSECRPFRELSEFSEMRARDPQFVVIO-NEDLG 1001
DB 354 KGERLPQPPICTIDVYIMVKWCMIDASRPKRELIAEFSEMRARDPQFVVIOQDERMH 413

QY 1002 PASPLDSTFYRSLLEDDMDGLDAEYLVPOQGFCCPDPAFGAGMVHRRHSSTRSG 1061
DB 414 LPSPTDSKFYRLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454

QY 1062 GGDLTLGLEPSEEEAPRSP-----APSEGAGSDVFDGLGMGAAGKGLQSLPHTDPSPLQ 1116
DB 455 -----SRTPLLSLSATSNNATNCIDRNGG-----H----- 481

QY 1117 RYSEDPVTPLPSETDGYVAPLTCSPQPEYVQPDVVRPQPPSPREGPLPAARPAAGAT-LER 1175
DB 482 -----PVREDGFL-----PAPEVNO--LMPKPFSTAMVQNIYVYISLTAISK 523

QY 1176 AKTLSPKNGVVKDVFAPFGGAVENPEYL 1203
DB 524 LPIDRYQN-----SHSTAVDNPYL 544
```

RESULT 14

GOFPE

epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Drosophila melanogaster
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C:Accession: A00640; A38021
R:livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A:Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
A:Reference number: A00640; MUID:85124611; PMID:2982499

A:Accession: A00640
A:Molecule type: DNA
A:Residues: 1-1330 <LIV>
A:Cross-references: EMBL:K03054
R:Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A:Title: A drosophila genomic sequence with homology to human epidermal growth factor re
A:Reference number: A38021; MUID:85137938; PMID:2983232
A:Accession: A38021
A:Molecule type: DNA
A:Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK' <WAD>
A:Cross-references: EMBL:X02293; NID:g7922; PIDN:CAA26157.1; PID:g929565
C:Comment: This sequence is tentative because the introns have not been identified.
C:Genetics:
A:Gene: FlyBase:Egfr
A:Cross-references: FlyBase:FBgn0003731
A:Map position: 2 57F
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP, autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph
F:1-732/Domain: extracellular #status predicted <EXT>
F:733-764/Domain: transmembrane #status predicted <TMM>
F:765-1330/Domain: intracellular #status predicted <INT>
F:808-1072/Domain: protein kinase homology <SKIN>
F:816-824/Region: protein kinase ATP-binding motif
F:122-300, 324, 363, 518, 688, 695, 700/Binding site: carbohydrate (Asn) (covalent) #status pr
F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:843/Active site: Lys #status predicted
F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predict

| | Query Match | Best Local Similarity | 24.0%; Score 1643.5; DB 1; Length 1330; |
|----|--------------|---|---|
| | Matches 412; | Conservative 181; | Mismatches 415; Indels 377; Gaps 39; |
| Qy | 80 | VOGYVLIAHNVQRVQLRLIRIVRGTOLF-----EDNYALAVLDNGDPLNNTTPTVTGASP | 134 |
| Db | 38 | ITNYIVIGLDLPCTLSYRLQIIRGRTLFLSLVVEEKYALFV-----TY | 81 |
| Qy | 135 | GGRLQLRSLTEILAGGVLIQRNPOLCYODTILWKDIFHKNNQLALTLIDNRSRACHP | 194 |
| Db | 82 | SKMYTLEIPDLRDVLNGQGVGHNNYNLCHMRTIQMSEIVSNGTDAYNYDFTAPERCPK | 141 |
| Qy | 195 | CSPMCKGSRGWGESSDCQSLTRTVACGGCA--RCKGPLPTDDCCHEQCAAGCTGPKHSDC | 252 |
| Db | 142 | CHESCTHG-CWGEKPKCKFKSLTCSQPCAGGRCYCPKPRECCHLFCAGCGTGTQKDC | 200 |
| Qy | 253 | LACLPHNHSIGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACPNYLSLTDVGSCT | 312 |
| Db | 201 | IACKNFFDEAVSKEECPMPKRYNPTTYVLETPNEGYAYGATCVKECP-GHLLRDNGACV | 259 |
| Qy | 313 | LVCPLHNQEVTAEDGTQRCEKSKPCARVCYGLGMFNNFTVSFWLRPKVSASHLEEFAG | 372 |
| Db | 260 | RSCPOQMDKGE-----CVPCKNGPCPKTCF-----VTVLHAGNIDSPRN | 300 |
| Qy | 373 | CKKIFGSLAFLPESFDG--DPASNTA-----PLQEQLOQVFETLEEITGYLYTSAMPDS | 424 |
| Db | 301 | CTVIDGNIRILOQTFSGFDQVYANYTMGPRIPLDPERREVFTSVKEITGYLNTGTHPQ | 360 |
| Qy | 425 | LPDLSVFQNLQVIRGRILHNGAY-SLTQGLGTSWLGLRSLRELGSGLALIHNNTHLCFV | 483 |
| Db | 361 | FRNLSYFRNLETHIGRQLMESMAALAVKSSLSYLEMRNLKQISSGVSIVQHNRDLGV | 420 |
| Qy | 484 | HTVPMDOLFPRNPHOALLHTANRPEDEC----- | 510 |
| Db | 421 | SNIRPATAQKEPEQKVWVNNENLEADLCGKFLTILISVQHNIIMHIFAIKREKNHLLGSV | 480 |
| Qy | 511 | ----- | 510 |
| Db | 481 | QRGRLLGSHGSPVYLOELQFOWHLHRLWLVIQVINSITQDKSNEHQLTDACYSPSVPT | 540 |
| Qy | 511 | ----- | 512 |
| Db | 541 | SLTIERARYAIQSAGLAMELEQITARSASNRHSKTLPAEGRQVPRWVFLGVCASARAGIA | 600 |

C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

| | | | | |
|-----------------------|--------|---|-----------------|---------------------|
| Query Match | 23.98; | Score 1640; | DB 2; | Length 545; |
| Best Local Similarity | 54.9%; | Pred. No. 3e-60; | | |
| Matches | 345; | Conservative 69; | Mismatches 122; | Indels 92; Gaps 15; |
| Qy | 584 | GPEADQCVACAHYKDPDFCVARCPGKVPDLSYMPIWKFPDEGACQPCPINCTHSCVDL | 643 | |
| Db | 1 | GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCOLCHPNCCTCKGKP | 57 | |
| Qy | 644 | DDKCPAEQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRLLOETEL | 702 | |
| Db | 58 | GLEGCP---NGSKTPSIAAGVWGGLCLVVVGLIGLYLRRR-HIVKRTLRRLQEREL | 113 | |
| Qy | 703 | VEPLTPSGAMPNQAHRIILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLR | 762 | |
| Db | 114 | VEPLTPSGEAPNQAHRIILKETEFKKVKVLGFGAFGVYKGLWIPGEKVTIPVAIKELR | 173 | |
| Qy | 763 | ENTSPKANKEILDEAVVMAGVSGPYVSRLLGICLTSTVOLVTOLMPYGCILLDHVRENRR | 822 | |
| Db | 174 | EATSPKANKEILDEAVVMASVDNPHVCRLLGICLTSTVOLITQIMPYGCLLDYIREHKDN | 233 | |
| Qy | 823 | LGSQDLLNMCQIAKGMVLEVDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE | 882 | |
| Db | 234 | IGSOYLLNMCVQIAKGMVLEERHLVHRDLAARNVLVKTQDVKITDFGLAKQLGADEKE | 293 | |
| Qy | 883 | YHADGKVPKIMWALESIILRRRTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIDPLE | 942 | |
| Db | 294 | YHAEGGKVPKIMWALESIILRIYTHOSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLE | 353 | |
| Qy | 943 | KGERLPOPPICTIDVYIMVYKMWIDSECRPRFRELVSFESRMARDPQRFVVIQ-NEDLG | 1001 | |
| Db | 354 | KGERLPOPPICTIDVYIMVYKMWSDADSRPKFRELIAEFKWARDPPRYLVIQGDERRH | 413 | |
| Qy | 1002 | PASPLDSTFYRSLLDDMDGLVDAAEYLVPOQGFPCDPAPGAGGMVHHRHSSSTRSG | 1061 | |
| Db | 414 | LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST---454 | | |
| Qy | 1062 | GGDLTLGLEPSEEAAPRSPL-----APSEGAGSDVFDGLGMGAAGKLSLPTHPDPSPLQ | 1116 | |
| Db | 455 | -----SRTPLLSLSATSNNSATNCIDRNGG-----H-----481 | | |
| Qy | 1117 | RYSEDPTVPLPSETDGVVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGAT-LER | 1175 | |
| Db | 482 | -----PVREDGFL-----PAPEYVQ--LMPKXFTAMVQNIYNYISLTAISK | 523 | |
| Qy | 1176 | AKTLSPGKNGVVKDVFAFGGAVENPEYL | 1203 | |
| Db | 524 | LPMDSRYQN-----SHSTAVDNPEYL | 544 | |

Search completed: July 22, 2003, 09:09:31
Job time : 31.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-103-117-12
Perfect score: 6809
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| 1 | 6730 | 98.8 | 1255 | 1 ERB2_HUMAN | P04626 homo sapien |
| 2 | 5963 | 87.6 | 1257 | 1 ERB2_RAT | P06494 rattus norv |
| 3 | 5941.5 | 87.3 | 1254 | 1 ERB2_MESAU | Q05053 mesocricetu |
| 4 | 3158 | 46.4 | 1210 | 1 EGFR_HUMAN | P00533 homo sapien |
| 5 | 3131 | 46.0 | 1210 | 1 EGFR_MOUSE | Q01279 mus musculu |
| 6 | 3003.5 | 44.1 | 1308 | 1 ERB4_HUMAN | Q15303 homo sapien |
| 7 | 2984 | 43.8 | 1308 | 1 ERB4_RAT | Q62956 rattus norv |
| 8 | 2715.5 | 39.9 | 1167 | 1 XMRK_XIPMA | P13388 xiphophorus |
| 9 | 2430.5 | 35.7 | 1342 | 1 ERB3_HUMAN | P21880 homo sapien |
| 10 | 2359.5 | 34.7 | 1339 | 1 ERB3_RAT | Q62799 rattus norv |
| 11 | 1976 | 29.0 | 1426 | 1 EGFR_DROME | P04412 drosophila |
| 12 | 1749.5 | 25.7 | 634 | 1 ERBB_ALV | P00534 avian leuko |
| 13 | 1703 | 25.0 | 604 | 1 ERBB_AVIER | P00535 avian eryth |
| 14 | 1630 | 23.9 | 540 | 1 ERBB_AVIEU | P11273 avian eryth |
| 15 | 1614 | 23.7 | 703 | 1 EGFR_CHICK | P13387 gallus gall |
| 16 | 1311 | 19.3 | 1323 | 1 LT23_CABEL | P24348 caenorhabdi |
| 17 | 1142.5 | 16.8 | 245 | 1 ERB2_MOUSE | P70424 mus musculu |
| 18 | 733 | 10.8 | 1363 | 1 ILPR_BRALA | O02466 branchiosto |
| 19 | 718 | 10.5 | 1382 | 1 INSR_HUMAN | P06213 homo sapien |
| 20 | 713 | 10.5 | 1383 | 1 INSR_RAT | P06213 homo sapien |
| 21 | 712.5 | 10.5 | 1372 | 1 INSR_MOUSE | P15127 rattus norv |
| 22 | 699 | 10.3 | 1300 | 1 IRR_MOUSE | P15208 mus musculu |
| 23 | 694 | 10.2 | 1607 | 1 MIPR_LYNST | Q9wt14 mus musculu |
| 24 | 690 | 10.1 | 1297 | 1 IRR_HUMAN | Q25410 lymnaea sta |
| 25 | 685.5 | 10.1 | 1297 | 1 IRR_MOUSE | P14616 homo sapien |
| 26 | 684 | 10.0 | 1477 | 1 HTK7_HYDAT | P14617 cavia porce |
| 27 | 651 | 9.6 | 1367 | 1 IGIR_HUMAN | Q25197 hydra atten |
| 28 | 644 | 9.5 | 1390 | 1 INSR_ADAE | P08089 homo sapien |
| 29 | 641 | 9.4 | 1373 | 1 IGIR_MOUSE | Q93105 aedes aegypt |
| 30 | 637.5 | 9.4 | 1370 | 1 IGIR_RAT | P60751 mus musculu |
| 31 | 617 | 9.1 | 2146 | 1 INSR_DROME | P24062 rattus norv |
| 32 | 605 | 8.9 | 987 | 1 EPB4_HUMAN | P09208 drosophila |
| 33 | 591.5 | 8.7 | 984 | 1 EPB1_CHICK | P54760 homo sapien |
| | | | | | Q07494 gallus gall |

ALIGNMENTS

RESULT 1

ERB2_HUMAN

| | | | | |
|----|--|-----------|------|----------|
| ID | ERB2_HUMAN | STANDARD; | PRT; | 1255 AA. |
| AC | P04626; | | | |
| DT | 13-AUG-1987 (Rel. 05, Created) | | | |
| DT | 13-AUG-1987 (Rel. 05, Last sequence update) | | | |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update) | | | |
| DE | Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112) | | | |
| DE | (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell | | | |
| DE | surface receptor HER2) (MLN 19). | | | |
| GN | ERBB2 OR HER2 OR NGI OR NEU. | | | |
| OS | Homo sapiens (Human) | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | |
| OX | NCBI_TaxID=9606; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=86118663; PubMed=3003577; | | | |
| RA | Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N., | | | |
| RA | Saito T., Toyoshima K.; | | | |
| RT | "Similarity of protein encoded by the human c-erbB-2 gene to | | | |
| RT | epidermal growth factor receptor."; | | | |
| RL | Nature 319:230-234 (1986). | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=86070181; PubMed=2999974; | | | |
| RA | Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A., | | | |
| RA | McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J., | | | |
| RA | Francce U., Levinson A., Ullrich A.; | | | |
| RT | "Tyrosine kinase receptor with extensive homology to EGF receptor | | | |
| RT | shares chromosomal location with neu oncogene."; | | | |
| RL | Science 230:1132-1139 (1985). | | | |
| RN | [3] | | | |
| RP | SEQUENCE OF 737-1031 FROM N.A. | | | |
| RX | MEDLINE=86016729; PubMed=2995967; | | | |
| RA | Semba K., Kanata N., Toyoshima K., Yamamoto T.; | | | |
| RT | "A v-erbB-related protooncogene, c-erbB-2, is distinct from the | | | |
| RT | c-erbB-1/epidermal growth factor-receptor gene and is amplified in a | | | |
| RT | human salivary gland adenocarcinoma."; | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985). | | | |
| RN | [4] | | | |
| RP | VARIANTS VAL-654 AND VAL-655. | | | |
| RX | MEDLINE=93194196; PubMed=8095488; | | | |
| RA | Ehsani A., Low J., Wallace R.B., Wu A.M.; | | | |
| RT | "Characterization of a new allele of the human ERBB2 gene by allele- | | | |
| RT | specific competition hybridization."; | | | |
| RL | Genomics 15:426-429 (1993). | | | |
| CC | -I- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, | | | |
| CC | ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A | | | |
| CC | POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF- | | | |
| CC | ALPHA AND AMPHIREGULIN. | | | |
| CC | -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein | | | |
| CC | tyrosine phosphate. | | | |
| CC | -I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS | | | |
| CC | (POTENTIAL). | | | |
| CC | -I- SUBCELLULAR LOCATION: Type I membrane protein. | | | |

Q03145 mus musculu
P07949 homo sapien
P54761 mus musculu
P29117 homo sapien
P09759 rattus norv
Q91736 xenopus lae
P54762 homo sapien
Q91571 xenopus lae
Q00944 gallus gall
Q91738 xenopus lae
P53356 hydra atten
P34152 mus musculu

| | | | |
|----------|---|--|---------------|
| Qy | 721 | RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILD EAYVMAGV | 780 |
| Db | 721 | RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILD EAYVMAGV | 780 |
| Qy | 781 | YVSRLLGICLTSTVOLTVQLMPPYGCCLLDHVRENRGRIGSODLNCWQOIAKNGSYL | 840 |
| Db | 781 | YVSRLLGICLTSTVOLTVQLMPPYGCCLLDHVRENRGRIGSODLNCWQOIAKNGSYL | 840 |
| Qy | 841 | LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPICKWMALESILRR | 900 |
| Db | 841 | LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPICKWMALESILRR | 900 |
| Qy | 901 | HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIM | 960 |
| Db | 901 | HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIM | 960 |
| Qy | 961 | IDSECRPRFRELVESEFRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM | 1020 |
| Db | 961 | IDSECRPRFRELVESEFRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM | 1020 |
| Qy | 1021 | EEYLVPOQGFCCPDPAAGGVMVHRRSSSTRSGGDLTLGLPESEEEAPRSP | 1080 |
| Db | 1021 | EEYLVPOQGFCCPDPAAGGVMVHRRSSSTRSGGDLTLGLPESEEEAPRSP | 1080 |
| Qy | 1081 | AGSDVDFDGLGMGAAGKGLSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPL | 1140 |
| Db | 1081 | AGSDVDFDGLGMGAAGKGLSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPL | 1140 |
| Qy | 1141 | NQPDVRPOPSPREGPLPAARPAAGATLIERAKTILSPGNKVGVKDFAFGGAVEN | 1200 |
| Db | 1141 | NQPDVRPOPSPREGPLPAARPAAGATLIERAKTILSPGNKVGVKDFAFGGAVEN | 1200 |
| Qy | 1201 | GGAAPQHPHPAPSPAFDNLVYDQDPPERGAAPPSTFKGTPTAENPEYLG | 1255 |
| Db | 1201 | GGAAPQHPHPAPSPAFDNLVYDQDPPERGAAPPSTFKGTPTAENPEYLG | 1255 |
| RESULT 2 | | | |
| ERB2 | RAT | STANDARD; | PRT; 1257 AA. |
| AC | P06456; | | |
| DT | 01-JAN-1988 (Rel. 06, Created) | | |
| DT | 15-DEC-1998 (Rel. 37, Last sequence update) | | |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update) | | |
| DE | Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112) | | |
| DE | (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor | | |
| DE | receptor-related protein). | | |
| GN | ERBB2 OR NEU. | | |
| OS | Rattus norvegicus (Rat). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Rattus. | | |
| NCBI | TaxID=10116; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TISSUE=Neuroblastoma; | | |
| RA | MEDLINE=86118662; PubMed=3945311; | | |
| RA | Bargmann C.I., Hung M.-C., Weinberg R.A.; | | |
| RT | "The new oncogene encodes an epidermal growth factor receptor-related | | |
| RT | protein."; | | |
| RL | Nature 319:226-230(1986). | | |
| RP | [2] | | |
| RP | SEQUENCE OF 852-905 FROM N.A. | | |
| RC | TISSUE=Sciatic nerve; | | |
| RC | MEDLINE=91222560; PubMed=2025425; | | |
| RA | Lai C., Lemke G.; | | |
| RT | "An extended family of protein-tyrosine kinase genes differentially | | |
| RT | expressed in the vertebrate nervous system."; | | |
| RL | Neuron 6:691-704(1991). | | |
| RN | [3] | | |
| RP | STRUCTURE BY NMR OF 650-668. | | |
| RP | MEDLINE=92155181; PubMed=1346763; | | |
| RA | Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D., | | |

```
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 87.6%; Score 5963; DB 1; Length 1257;
Best Local Similarity 87.4%; Pred. No. 9.3e-313;
Matches 1099; Conservative 51; Mismatches 105; Indels 2; Gaps 2;

Qy 1 MELAAALCRWGLLLALLPPGAASCTVCTGDMKRLPASPETHDMLRHLHYOGCVVQGNL 60
Db 1 MELAAALCRWGLLLALLPPGIAGTQVCTGDMKRLPASPETHDMLRHLHYOGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQYVLIAHNQVQPLQRLIRVGTQQLFEDNYALAVLDNG 120
Db 61 ELTYVPANASLSFLQDIOEVQYVLIAHNQVQPLQRLIRVGTQQLFEDKYALAVLDNR 120
Qy 121 DPLNNQYIKANSKF-IGITELQLRLSLTEILKGGVLIQBNPOLCYODTILWKDIFHKNNQL 179
Db 121 DPQDNVAASTPGRTEGLRELQLRLSLTEILKGGVLIIRGNPOLCYQDMVLWKDVFRRKNNQL 180
Qy 180 ALTLIDNRSRACHPCSPMKGSCWGESSEDCSLTRTVCAAGCARCKGPLTDCCHEQ 239
Db 181 APVDIDNRSRACHPCAPACKDNHCWGESPEDCQLTGTCTSCARCKGPLTDCCHEQ 240
Qy 240 CAAGCTGKHSDDLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTAC 299
Db 241 CAAGCTGKHSDDLACLFHNSGICELHCPALVYNTDTFESMENPEGRTYFGASCVTTC 300
Qy 300 PYNLYSTDVGSCTLVCPLNHOEVAEDGTORCEKSPCARVCYGLGMEHLRVRVATSA 359
Db 301 PYNLYSTEVGSCTLVCPNNOEVAEDGTORCEKSPCARVCYGLGMEHLRGARAITSD 360
Qy 360 NIOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLQVPTLEEITGYLIYISAWPDSL 419
Db 361 NVQEFDCCKKIFGSLAFIPESFDGDPSSGIAPLPEQLQVPTLEEITGYLIYISAWPDSL 420
Qy 420 PDLVSFQNLQVIRIILHNGAYSILTLOGLSISWGLSLRLSRLGSLALIHHTHLCFVHT 479
Db 421 RDLVSFQNLRIIRIILHDGAYSILTLOGLSISLGLSLRLSRLGSLALIHNAHLCFVHT 480
Qy 480 VPMDFLFRNPHOALLHTANREDE-CYEGELACHQLCARHCWGPGTQCNCSCQFLRGQ 538
Db 481 VPMDFLFRNPHOALLHSGNREDEULCVSSGLVCNSLCAHGCWGPPTQCNCSCQFLRGQ 540
Qy 539 ECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCA 598
Db 541 ECVEECRVWGLPREYVSDKRLPCHPECPQNSSETCFGEADQCAAHYKDKSSCVA 600
Qy 599 RCPGSGVPDLSYMPIWKPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIVSAV 658
Db 601 RCPGSGVPDLSYMPIWKYPDEEGICQPCPINCTHSCVDLDERGCPAQASPLTSIVATV 660
Qy 659 VGIILVVVGLVGVFGLIKRQOKIRKYMRLLOETELVEPLTSGAMPQAQWRIKET 718
Db 661 VGIILVVVGLVGVFGLIKRQOKIRKYMRLLOETELVEPLTSGAMPQAQWRIKET 720
Qy 719 ELRKVKVLSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAVG 778
Db 721 ELRKVKVLSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAVG 780
Qy 779..SPVYVRLGLICLTSTVQLTQMPYGCCLLDHVRNRRGLSGQDLLNMCQIAKGMYSYLED 838
```

```
Db 781 SPVYVRLGLICLTSTVQLTQMPYGCCLLDHVRHGRGLSGQDLLNMCQIAKGMYSYLED 840
Qy 839 VRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVPYIKWMALESILRRR 898
Db 841 VRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVPYIKWMALESILRRR 900
Qy 899 FTHOSDVSQVYVWELMTFCAPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMWKC 958
Db 901 FTHOSDVSQVYVWELMTFCAPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMWKC 960
Qy 959 WMIDSECRPRELVSFSEFMRARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDDMGDLV 1018
Db 961 WMIDSECRPRELVSFSEFMRARDPQRFVVIQNEDELGPSSPMDSTFYRSLLEDDDMGDLV 1020
Qy 1019 DAESYLVPOQFFCFDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPSEEAAPRPLAPS 1078
Db 1021 DAESYLVPOQFFCFDPPTGPTGTAHRRHRSSTSRSGGDLTLGLEPSEEGPRPLAPS 1080
Qy 1079 EGAGSDVFDGDLGMAAKGLQLPHTDPSPLQRYSEDPVLPSETDGVVAPLTCSPOPE 1138
Db 1081 EGAGSDVFDGDLGMAVTKGLQSLPSHLSPQRYSEDPVLPSETDGVVAPLTCSPOPE 1140
Qy 1139 YVNPQDVRPOPSPREGPLPAARPAAGATLERAKTLPCKNGVVKDVFAFGGAVENPEYLT 1198
Db 1141 YVNPQEQVQPPPLTPEGPLPPVRPAGATLERPKTLPCKNGVVKDVFAFGGAVENPEYLV 1200
Qy 1199 PQGAAAPQHPPPAFSPAFNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
Db 1201 PREGTASPPHPSPAFSPAFNLVYWDQNSSEQPPSPNEFCTPTAENPEYLGLDVVPV 1257

RESULT 3
ERB2_MESAU
ID ERB2_MESAU STANDARD; PRT; 1254 AA.
AC Q60533;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) - (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PFM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

| | | | | | |
|-----|---|----|------|--|------|
| CC | EMBL; D16295; BAA03801.1; . | QY | 61 | ELTYLPTNASLFLQDIQEVQGYVLIHQNVRQVPLQRLRIVRGTLQFEDNYALAVLDNG | 120 |
| CC | HSSP; P11362; 1FGK. | Db | 61 | ELTYLPANATLSFLQDIQEVQGYVLIHQNVRQVPLQRLRIVRGTLQFEDNYALAVLDNR | 120 |
| DR | InterPro; IPR000494; EGFR_L domain. | QY | 121 | DPLNNQYIKANSKFTIGITELQLRSUTEILKGGVLIQORNPOLCYQDITILWKDIPKHNQLA | 180 |
| DR | InterPro; IPR000719; Euk_pKinase. | Db | 121 | DPLDNNVTATGRTPEGLRELQLSUTEILKGGVLIQORNPOLCYQDITILWKDIPKHNQLA | 180 |
| DR | InterPro; IPR002174; Furin-like. | QY | 181 | LTLIDNRSRACHPCS PMCKSGRCWGESSEDCQSLTRTVACGCGCARCKGPLPTDCCHEQC | 240 |
| DR | InterPro; IPR004019; YLP motif. | Db | 181 | PVDIITNRSRACHPCS PMCKSGRCWGESSEDCQSLTRTVACGCGCARCKGPLPTDCCHEQC | 240 |
| DR | Pfam; PF00069; pkinase; 1. | QY | 241 | AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP | 300 |
| DR | Pfam; PF00757; Furin-like; 1. | Db | 241 | AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP | 300 |
| DR | Pfam; PF01030; Recep_L_domain; 2. | QY | 301 | YNYLSTDVSGCTLVCPLNHNEVTABDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN | 360 |
| DR | Pfam; PF02757; YLP; 2_domain; 1. | Db | 301 | YNYLSTDVSGCTLVCPLNHNEVTABDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN | 360 |
| DR | ProDom; PD000001; Euk_pkinase; 1. | QY | 361 | IOEPAGCKKIFGSLAFLESFDPGDPASNTAPLOEQLOVFETLEETGYLYISAMPDSL | 420 |
| DR | SMART; SM00261; FU; 3. | Db | 361 | IOEPAGCKKIFGSLAFLESFDPGDPASNTAPLOEQLOVFETLEETGYLYISAMPDSL | 420 |
| DR | SMART; SM00219; TyrKc; 1. | QY | 421 | DLVSFQNLQVIRGRILHNGAYSLTLQGLGISMLGRSLRELGLALIHNNTHLCFVHTV | 480 |
| DR | PROSITE; PS00107; PROTEIN KINASE ATP; 1. | Db | 421 | DLVSFQNLQVIRGRILHNGAYSLTLQGLGISMLGRSLRELGLALIHNNTHLCFVHTV | 480 |
| DR | PROSITE; PS00109; PROTEIN_KINASE_TYR; 1. | QY | 481 | PDQLFRNPHOALLHNTANRDECEVGEGLACHOLCARHCWGPGPTQCVCNCSHFLRGQEC | 540 |
| DR | PROSITE; PS00111; PROTEIN_KINASE_DOM; 1. | Db | 481 | PDQLFRNPHOALLHNTANRDECEVGEGLACHOLCARHCWGPGPTQCVCNCSHFLRGQEC | 540 |
| KW | Transmembrane; Glycoprotein; Multigene family; Receptor; Signal; | QY | 541 | VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFCPEADOCVACAHYKOPPCVARG | 600 |
| KW | Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation; | Db | 541 | VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFCPEADOCVACAHYKOPPCVARG | 600 |
| KW | Proto-oncogene; Disease mutation. | QY | 601 | PSGVKPDLSYMPKPFDEEGACQPCPINCTHSCVLDLDDKGGCPAEQASPLTSIVSAVVG | 660 |
| FT | SIGNAL 1 21 POTENTIAL. | Db | 601 | PSGVKPDLSYMPKPFDEEGACQPCPINCTHSCVLDLDDKGGCPAEQASPLTSIVSAVVG | 660 |
| FT | CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2. | QY | 661 | ILLVVVLGVVGIILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL | 720 |
| FT | DOMAIN 22 652 EXTRACELLULAR (POTENTIAL). | Db | 661 | ILLVVVLGVVGIILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL | 720 |
| FT | DOMAIN 653 675 POTENTIAL. | QY | 721 | RKVVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGLSP | 780 |
| FT | DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL). | Db | 721 | RKVVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGLSP | 780 |
| FT | DOMAIN 158 368 CYS-RICH. | QY | 781 | YVSRLLGICLTSTVQLVTQLMPYGCCLLDHVRNRCGLSQDILLNWCQIAKGMVLEEDVR | 840 |
| FT | DOMAIN 472 644 CYS-RICH. | Db | 781 | YVSRLLGICLTSTVQLVTQLMPYGCCLLDHVRNRCGLSQDILLNWCQIAKGMVLEEDVR | 840 |
| FT | DOMAIN 720 987 PROTEIN KINASE. | QY | 841 | LVRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPPIKMALESILRRRFT | 900 |
| FT | NP_BIND 726 734 ATP (BY SIMILARITY). | Db | 841 | LVRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPPIKMALESILRRRFT | 900 |
| FT | BINDING 753 753 ATP (BY SIMILARITY). | QY | 901 | HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPDPPICITIDVYMIWKCM | 960 |
| FT | ACT_SITE 845 845 BY SIMILARITY. | Db | 901 | HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPDPPICITIDVYMIWKCM | 960 |
| FT | DISULFID 195 204 BY SIMILARITY. | QY | 961 | IDSECRPRFRELVSSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA | 1020 |
| FT | DISULFID 199 212 BY SIMILARITY. | Db | 961 | IDSECRPRFRELVSSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA | 1020 |
| FT | DISULFID 216 244 BY SIMILARITY. | QY | 1021 | EYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLPESSEAPRSLAPSEG | 1080 |
| FT | DISULFID 240 252 BY SIMILARITY. | Db | 1021 | EYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLPESSEAPRSLAPSEG | 1080 |
| FT | DISULFID 255 264 BY SIMILARITY. | QY | 1081 | AGSDVFDGDLGMGAAGKGLQSLPHTDPPSQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV | 1140 |
| FT | DISULFID 268 295 BY SIMILARITY. | Db | 1081 | AGSDVFDGDLGMGAAGKGLQSLPHTDPPSQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV | 1140 |
| FT | DISULFID 299 311 BY SIMILARITY. | QY | | | |
| FT | DISULFID 315 331 BY SIMILARITY. | Db | | | |
| FT | DISULFID 334 338 BY SIMILARITY. | QY | | | |
| FT | DISULFID 511 520 BY SIMILARITY. | Db | | | |
| FT | DISULFID 515 528 BY SIMILARITY. | QY | | | |
| FT | DISULFID 531 540 BY SIMILARITY. | Db | | | |
| FT | DISULFID 544 560 BY SIMILARITY. | QY | | | |
| FT | DISULFID 563 576 BY SIMILARITY. | Db | | | |
| FT | DISULFID 567 584 BY SIMILARITY. | QY | | | |
| FT | DISULFID 587 596 BY SIMILARITY. | Db | | | |
| FT | DISULFID 600 623 BY SIMILARITY. | QY | | | |
| FT | DISULFID 626 634 BY SIMILARITY. | Db | | | |
| FT | DISULFID 630 642 BY SIMILARITY. | QY | | | |
| FT | MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY). | Db | | | |
| FT | MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY). | QY | | | |
| FT | CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL). | Db | | | |
| FT | CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL). | QY | | | |
| FT | CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL). | Db | | | |
| FT | CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL). | QY | | | |
| FT | CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL). | Db | | | |
| FT | CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL). | QY | | | |
| FT | CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL). | Db | | | |
| FT | VARIANT 658 658 V -> E (IN ONCOGENIC NEU). | QY | | | |
| FT | VARIANT 659 659 V -> E (IN ONCOGENIC NEU). | Db | | | |
| SEQ | SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64; | QY | | | |

Query Match
 Best Local Similarity 87.3%; Score 5941.5; DB 1; Length 1254;
 Matches 1093; Conservative 58; Mismatches 103; Indels 1; Gaps 1;
 1 MELAAACRWGLLLALLPPGAASGTQCTGTDMLRASPETHLDMLRHLRYGCGVQGNL 60
 1 MELAAACRWGLLLALLSPGASGTQCTGTDMLRASPETHLDIVRHLRYGCGVQGNL 60


```

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
      (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 6).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

Query Match 46.0%; Score 3131; DB 1; Length 1210;
Best Local Similarity 49.5%; Pred. No. 8.9e-161;
Matches 630; Conservative 171; Mismatches 361; Indels 110; Gaps 23;

Qy 11 LLLALLPPGAA--STQVCTGTDMLKRLPASPEHLDMLRLHLYQGVQVQGNLELYLPTN 68
Db 14 LTLCAAGGALAEKKVCGQTSNRUTQGTGFEDHFLSQRMYNCEVVLGNLEITYQVRN 73
Qy 69 ASLSFLDIOBVOGYVLIHQAQVQVPLQRLIRVGTQTFEDNYALAVLDNGDPLNNQV1 128
Db 74 YDLSFLKTIQEVAGVLIHQAQVQVPLQRLIRVGTQTFEDNYALAVLDNGDPLNNQV1 128
Qy 129 KANSKFGITELQRLSLTEILKGGVLIQORNQOLCQVDITLWKDI-----FHKNQALTLI 184
Db 125 -YGNRTGLRLPRLNRLQEIILIGAVRFNSNFIILCNMDTIQWRDVIQVNFMSMDL--- 180
Qy 185 DTRNSRACHPCS PMCKGSRGWGSESSDCQSILTRITVCAGGCA-RCKGPLPTDCHEQCAAG 243
Db 181 -QSHPSSCPCKDPCPNCSGCHGGEENCKLTKIICAQCSHRCKRGRSPSCCHNQCAAG 239
Qy 244 CTGPKHSCDLACHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNY 303
Db 240 CTGPRESCLVCQKQFQDEATCKDTCPLMLNPTTYQMDVNPPEGKYSFGATCVKCPNY 299
Qy 304 LSTDVGSCTLCPLHNVQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSANTQE 363
Db 300 VVTDHGSVCRAAGPDYEV-BEDGIRKCKKCGPCRCVNGIGIGEFKDKTILSINATNIK 358
Qy 364 FAGCKIFGSLAFIPESFDGDPASNTAPLOEQVFTLEITCYLVISAWPDSLPLS 423
Db 359 FKYTAISGDHLIIPVAFKGSFTRTPDPRELEILKTVREITGFLLIQAWPNDTDLH 418
Qy 424 VFQNLQVIRGLHNGAYSLTLOGLIGISWLGRLSRLSELGSLALITHNTHLCFVHTVPWD 483
Db 419 AFENLEIIRGTRKQHQPSLAVVGLNITSGLRSLKEISDGDVITSGRNLCYANTINWK 478
Qy 484 QLFNPHQALLHTANRDECEVGEGLACHQICARGHCWGPPTQCVNCSQFLRQECVEE 543
Db 479 KLFCTPNQTKIMNRAEKDKAVNVHVCNPLCSSEGCWGPEDPCVSCQVSRGECVEK 538
Qy 544 CRVLQGLPREVVARHCLPCHPEQOPNGSVTCFGEADQCVACAHYKDPFPCVAPCSG 603
Db 539 CNILGEPREFEVSECIQCHPECLPQMNITCTGRGPDNCIQCAHYIDGPHCVKTCFAG 598
Qy 604 VKPDLSYMPIWKFPEEGACQCPINCTHSCVDLDDKCPAEQASPLTISVAVGILL 663
Db 599 IMGENNTL-VNKYADANNVCHLCHANCTYGCAGPGLQCEVWPSPGKIPSTATGIVGGL 657
Qy 664 VVLGVVFGI-LIKEROOKIRKTYMRLLOTELVEPLTPSGAMPNQOMRILKETELRK 722
Db 658 FIVV-VALGIGLFWRRRHIRKTRLLRLQRELVEPLTPSGEAPNQAHRLKETEFKK 716
Qy 723 VKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKILDEAYVMAGVGPYV 782

```

```

Db 717 IKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKILDEAYVMASVDNPHV 776
Qy 783 SRLGICITSTVQLVLTQMPYGCILLDHYVRENRLGRSGODLLNWCQIAKGMYSLEDRVLV 842
Db 777 CRLGICITSTVQLITQMPYGCILLDHYVREKDNIGSQYLLNWCQIAKGMYSLEDRVLV 836
Qy 843 HRDLAARNVLKSPNHVKITDFGLARLDIDTEYHADGKVPYIKWMALESILRRFTTHQ 902
Db 837 HRDLAARNVLKTPQHVKITDFGLAKLLGAEKEYHAEGKVPYIKWMALESILHRYTHQ 896
Qy 903 SDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEGERLPQPPICITIDVYMWKCMID 962
Db 897 SDVWSYGVTVWELMTFGSKPYDGIIPASDISISLEGERLPQPPICITIDVYMWKCMID 956
Qy 963 SECPRFRELVSFESRMARDPQRFVIO-NEDLGASPLDSTFYRSLLDDDDMDGLVDAAE 1021
Db 957 ADSRPFRELILFESKWARDPQRYLVIQDGRMHLPSPDTSNFYALMDEEDMDVVDAD 1016
Qy 1022 EYLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLSEPEEEAPRSPAPSEGA 1081
Db 1017 EYLVPQOQFF-----NSPST-----SRTPLLSLSA 1042
Qy 1082 GSDVPDGLGMAKGLQSLPHTDPSPLORYSEDDTVPLPSET--DGYVAPLTCSPQPEY 1139
Db 1043 TSN-----NSTVACINRNGSCRVKEDAFQRYSSDPTGAVTEDNIDDAFL-----PVPEY 1092
Qy 1140 VNQPDVRPQPPSPREGPLPAARAGATLERAKTLPKNGVVKVDFAFGGAIVENPEYL-T 1198
Db 1093 VNQ-SVPRKPSGVQNPVYHNPQLHP-----AFGRDLHYQN--PHSNVGNPEYLT 1141
Qy 1199 POGGAAPQHPHPPAPSPAFDNLVYWDQ-----DP-----PERGAPSTPKGTPT 1242
Db 1142 AQ-----PTCLSSGFNSPALVIQKSHQMSLDNDPDYQDFFPKTKNGIKPG-PT 1191
Qy 1243 AENPEYGLDVP 1254
Db 1192 AENAEYLRVAPP 1203

RESULT 6
ID_ERB4_HUMAN STANDARD; PRT; 1308 AA.
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (tyrosine kinase-type cell surface receptor HER4).
GN ERB4 OR HER4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culoucou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family."
RN Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
[2]
RP SEQUENCE FROM N.A. (ISOFORM JM-B).
RC TISSUE=Fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Elenius K., Corfas G., Paul S., Choi C.-J., Rio C., Plowman G.D.,
RA Klagsbrun M.;
RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
RT tissue distribution and differential processing in response to
RT phorbol ester."
RN J. Biol. Chem. 272:26761-26768(1997).

```



```

Db      536 CNUYDGEFREFENGSCVCECDPOCEKMEGDGLLTCHGPGPNCTKCSHFKDPGNCVKEKCPD 595
QY      603 GVKPDLISYPIWKFPDEEGACQPCINCTHSCVDLDDKGC-----PAEQRASPL 651
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      596 GLOGANSF--IFKYADPDRECHPCNPCTGCGNGPTSHDCIYYPWTCHSTLPQOHAR--TPL 652
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      652 TSIVSAVV--GILLVVVLGVVGLILIKRQOKIRKYTHRRLLQETELVEPLTPSGAMPNOA 710
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      653 --IAAGVIGGLFLVIVGLTFVAVYVRKSIK-KKRALRREL-ETELVEPLTPSGTAPNOA 708
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      711 OMRIKETELRKVKVLGSGAFGVYKGIWIPDGENVKI PVAIKVLRENTSPKANKETLDE 770
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      709 QLRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKI PVAIKILNETTGPKANVFMD 768
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      771 AYVMAGVGSYPVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGRLSQDLLNWCQIA 830
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      769 ALIMASMDPHVRLVLLGVCLSPITQLVTQLMPHGCLLEYVHEHKDNIGSQLLNWCQIA 828
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      831 KGSYLEDVRLVHRDLAARVLKSPNVKIDTFLGLARLLDIDETEHADGGKVPKIKMA 890
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      829 KGMYLEERLRLVHRDLAARVLKSPNVKIDTFLGLARLLDIDETEHADGGKVPKIKMA 888
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      891 LESILRRRTHQSDVMSYGVTVVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTID 950
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      889 LECIHYRKTHQSDVMSYGVTVVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTID 948
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      951 VYMIMVKCWMIDSECRPRFRELSEFSRMARDPQRFVVIQNED-LGPASPLDSTFYRSL 1009
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      949 VYVMVVKCWMIDADSRPKELAAEFSSRMARDPQRFVVIQNED-LGPASPLDSTFYRSL 1008
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      1010 EDDMDGLDVAEYLVYVQQGFCFDPAPGAGGVHHRSSSTRSGGDLTLGLPSEEE 1069
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1009 DEEDLEMDMAEYLVYV-QAFNIPPP-----TYTSRAIRDSNRS-----EIGHSPPPAY 1056
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      1070 APRS-----FLAP--SEGAGSDVFDGDLGMAAGKGLQS 1100
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1057 TPMSGNQFYVRDGGFAAEQGVSVYPRAPTSTIEAPVAQCATAEIFDDSCCNTLKRKPA 1116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      1101 LPHDPSPIQRYSDDPTVPLPS-----ETDGYVAPLTCSPQPEYVYNQDVPRPQPSPR 1153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1117 PHVQEDSSQRYSDPTVPAERSPRGELDEEGYMTMRDKPKQEVLYNPVE----- 1167
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      1154 EGPLPAARAGATLERAKTLSPQKGVKVDVAFAGAVENPEYLTQGGNAOPHPHPPA- 1212
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1168 ENPFVSR-----KNGDLQ-----ALDNPEYHNASG-----PPKAE 1199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      1213 -----FSPAFDNLVYWDQDPPERGA--PPSTF 1237
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1200 DEVVNEPLVNTFANTLGAEYLNKLNLSMPEKAKAFDNPDIWNHSLPRSTLQHPDYL 1259
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      1238 KGTPT-----AENPEYL 1249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1260 QEYSTKYFYKQNGRIRPPIVAENPEYL 1285
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
ERB4_RAT
ID ERB4_RAT STANDARD; PRT; 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (BC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,

```

```

RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; AF041838; AAD08899.1; -.
CC EMBL; U52531; AAC53051.1; -.
CC HSSP; P11362; 1FGK.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_Pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP_motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP_2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 4.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
CC SIGNAL 1 25 POTENTIAL.
CC CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
CC DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 652 675 POTENTIAL.
CC DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).

```

| | | | | | |
|---|----------|---|--|--|--|
| FT | DOMAIN | 186 | 334 | CYS-RICH. | |
| FT | DOMAIN | 496 | 633 | CYS-RICH. | |
| FT | DOMAIN | 718 | 985 | PROTEIN KINASE. | |
| FT | NP BIND | 724 | 732 | ATP (BY SIMILARITY). | |
| FT | BINDING | 751 | 751 | ATP (BY SIMILARITY). | |
| FT | ACT_SITE | 843 | 843 | BY SIMILARITY. | |
| FT | DISULFID | 189 | 197 | BY SIMILARITY. | |
| FT | DISULFID | 193 | 205 | BY SIMILARITY. | |
| FT | DISULFID | 213 | 221 | BY SIMILARITY. | |
| FT | DISULFID | 217 | 229 | BY SIMILARITY. | |
| FT | DISULFID | 230 | 238 | BY SIMILARITY. | |
| FT | DISULFID | 234 | 246 | BY SIMILARITY. | |
| FT | DISULFID | 249 | 258 | BY SIMILARITY. | |
| FT | DISULFID | 262 | 289 | BY SIMILARITY. | |
| FT | DISULFID | 293 | 304 | BY SIMILARITY. | |
| FT | DISULFID | 308 | 323 | BY SIMILARITY. | |
| FT | DISULFID | 326 | 330 | BY SIMILARITY. | |
| FT | DISULFID | 503 | 512 | BY SIMILARITY. | |
| FT | DISULFID | 507 | 520 | BY SIMILARITY. | |
| FT | DISULFID | 523 | 532 | BY SIMILARITY. | |
| FT | DISULFID | 536 | 552 | BY SIMILARITY. | |
| FT | DISULFID | 555 | 569 | BY SIMILARITY. | |
| FT | DISULFID | 559 | 577 | BY SIMILARITY. | |
| FT | DISULFID | 580 | 589 | BY SIMILARITY. | |
| FT | DISULFID | 593 | 614 | BY SIMILARITY. | |
| FT | DISULFID | 617 | 625 | BY SIMILARITY. | |
| FT | DISULFID | 621 | 633 | BY SIMILARITY. | |
| FT | MOD_RES | 1162 | 1162 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). | |
| FT | MOD_RES | 1188 | 1188 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). | |
| FT | MOD_RES | 1258 | 1258 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). | |
| FT | MOD_RES | 1284 | 1284 | PHOSPHORYLATION (AUTO-) (BY SIMILARITY). | |
| FT | CARBOHYD | 138 | 138 | N-LINKED (GLCNAC. .) (POTENTIAL). | |
| FT | CARBOHYD | 174 | 174 | N-LINKED (GLCNAC. .) (POTENTIAL). | |
| FT | CARBOHYD | 253 | 253 | N-LINKED (GLCNAC. .) (POTENTIAL). | |
| FT | CARBOHYD | 358 | 358 | N-LINKED (GLCNAC. .) (POTENTIAL). | |
| FT | CARBOHYD | 410 | 410 | N-LINKED (GLCNAC. .) (POTENTIAL). | |
| FT | CARBOHYD | 473 | 473 | N-LINKED (GLCNAC. .) (POTENTIAL). | |
| FT | CARBOHYD | 495 | 495 | N-LINKED (GLCNAC. .) (POTENTIAL). | |
| FT | CARBOHYD | 548 | 548 | N-LINKED (GLCNAC. .) (POTENTIAL). | |
| FT | CARBOHYD | 576 | 576 | N-LINKED (GLCNAC. .) (POTENTIAL). | |
| FT | CARBOHYD | 620 | 620 | N-LINKED (GLCNAC. .) (POTENTIAL). | |
| FT | CONFLICT | 1062 | 1062 | S -> N (IN REF. 3). | |
| FT | CONFLICT | 1080 | 1082 | PYT -> SYR (IN REF. 3). | |
| FT | SEQUENCE | 1308 | 1308 | AA; D944BB0996A08B41 CRC64; | |
| Query Match 43.8%; Score 2984; DB 1; Length 1308; | | | | | |
| Best Local Similarity 45.4%; Pred. No. 7.6e-153; | | | | | |
| Matches 612; Conservative 192; Mismatches 383; Indels 160; Gaps 29; | | | | | |
| Qy | 1 | MELA-ALCRWGLL-- | ALLPPGAASQTQCTGDMKRLPASPETHLDMRLHLYQGCVVQ | 57 | |
| Db | 1 | MKLATGLWVGSLLVAARTVQPSASQSVCACTENKLSLSLEQOYRALRKYENCVV | 60 | | |
| Qy | 58 | GNLELTLPNTNASLFLQDTEOVGYVLIHNOVQVPLORLRIVRGTQJFEDNYALVL | 117 | | |
| Db | 61 | GNLEITSIEHNRDLFLRSREVGYVVALNQFRLPLENRLIIRGTLKYEDRYALAIF | 120 | | |
| Qy | 118 | DNGDPLNNQVYKANSKFIGITELQRLSLTEILKGGVLIORNPOLCYODTILWKDIFHKNN | 177 | | |
| Db | 121 | LN-----YKQGN-- | FGLEQLGLKNLTELINGGVVDQNKFLCYADTIHWQDIDVRNP | 171 | |
| Qy | 178 | QLALTLDITNRSRACHPCSPMKSGRCWGESSEDCQSLTRITVCAGGC-ARCKGLPTDCC | 236 | | |
| Db | 172 | PSNMTLVSTIGSSGCRCHKSCG-RCWGPTENHCQTLLTRTVCAEQCDGRCYGPVSDCC | 230 | | |
| Qy | 237 | HEOCAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMNPGRYTFGASCV | 296 | | |
| Db | 231 | HRCAGCGSGPKDTCDFACMNFNDSGACVTCQPTQFYNNPTTFOLEHFNKAYTYGAFV | 290 | | |
| Qy | 297 | TACPYNLYLTDVGSCTLVCPHLHNOEVTAEQGTQRCCKSPCARVCYGLGHEHLREVRV | 356 | | |
| Db | 291 | KKCPHFVY-VDSSSCVRCACFPSSKQEV--EENGKMKCECTDICPKACDGIQTGSLMSAQT | 348 | | |

RESULT 8
XMRK_XIPMA
ID XMRK_XIPMA
AC P13388;

STANDARD; PRT; 1167 AA.

| | | | |
|----|------|---|------|
| Qy | 357 | TSANIOEFAGCKKIFGSLAFLESFDPASNTAPLOEQVQVFETLEEITGYLVISAMP | 416 |
| Db | 349 | DSSNIDFKINCTKINGNLI FLVTGIHGDPPYNAIDAIPEKLVNRTVREITGFLNIQWP | 408 |
| Qy | 417 | DSLPLSVFQNLQVIRGRIHLHGAYSILTQGLGSIWGLRSLRELGSGLAIHHNTHICF | 476 |
| Db | 409 | PNMTDFSVFNLVTIGRVLYSGLSLILKQOGITSLQFQSLKETISAGNIVITONSLEY | 468 |
| Qy | 477 | VHTVPDOLRNPHOALLHTANRPEDECVGEGLACHOLCARGHGWGPGTQVNCVQFLR | 536 |
| Db | 469 | YHTINWTTFLSTVNQRIVRDNRRAENCTAGMYCNHLCSNDGCMGPGPDQCLSCRRFSR | 528 |
| Qy | 537 | GQECVEBCRVLQGLPREYVNAHCLPCHPEQCP--QNGSVTCFGEADOCVACAHYKDPFF | 595 |
| Db | 529 | GKICIESCNLYDGEFRFENGSI CVCDSQCEKMEGDLTCHGCPDNCTCKSHFGDPN | 588 |
| Qy | 596 | CVARCPGVPKDLSPYMPIMPKPDEEGACQPCPINCTHSCVDLDDKGC-----PA | 644 |
| Db | 589 | CVEKCPDVLQANSF--IFKYADQDRCHPCHPNCCTQCGNPTSHDCIYYPMTGHTLPQ | 646 |
| Qy | 645 | EQRASPLTSIVSAVV--GILLVVVLGVVFGIILIKRQOKIRKYTHRRLLQETELVEPLTPS | 703 |
| Db | 647 | HAR--TPL--IAAGVIGLFLVIMALTFVAVVRRKSIK--KKALRRFL--ETELVEPLTPS | 701 |
| Qy | 704 | GAMPNOAQMRLKETELRKVKVLCGSAFGTVYKGIWIPDGNVWIPVAIKVLRNTPSKA | 763 |
| Db | 702 | GTAPNOAQLRLKETELRKVKVLCGSAFGTVYKGIWIPDGNVWIPVAIKVLRNTPSKA | 761 |
| Qy | 764 | NKEILDYAVMAGVSPVSRLLGICLTSTVQLTQVLMYPYGLLDHVRNRLGSGDQL | 823 |
| Db | 762 | NVEFMDEALIMASVDHPLVRLLCVLSPTIQLVTLMPHGLCLLEYVHEHKDNIGSLL | 821 |
| Qy | 824 | NWCMOIAKGSYLEDVRLVHRDLAARNVVKSPNHVKITDGLARLLDIDETEVHADGCK | 883 |
| Db | 822 | NWCQIAKGMVLEERLVRHDLAARNVVKSPNHVKITDGLARLLDIDETEVHADGCK | 881 |
| Qy | 884 | VPIKWMALSLRRTFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQ | 943 |
| Db | 882 | MPIKWMALSLRRTFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQ | 941 |
| Qy | 944 | PPCTIDVYIMVKWMDSECRPRFRELSEFSRMDPDRFVVIQNEH--LGPASPLDS | 1002 |
| Db | 942 | PPCTIDVYIMVKWMDSECRPRFRELSEFSRMDPDRFVVIQNEH--LGPASPLDS | 1001 |
| Qy | 1003 | TFYRSLLEDDMDGLVDAAEYLVPOQGFCCDP-----APGA | 1039 |
| Db | 1002 | KFQNLLEDEDDMDAAEYLVLP--QAFNIPPIYTSRTRIDSNRSEIGHSPPPAYTPMS | 1060 |
| Qy | 1040 | GMVYHHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEGAGSDVFGDGLCMGAAGLQ | 1099 |
| Db | 1061 | GSQFVYQDGGFATQOG---MPMPYATATSTIPEAPVA--QCATAEFDDSCNGTLRKP | 1115 |
| Qy | 1100 | SLPHDPSLQRYSEDTVPLPS-----ETDGYVAPLTCSPQEVYVQPDVRPQPPSP | 1152 |
| Db | 1116 | VPHVEDSSTQRYSDPTVFAPERNAELDEEGYMTPMHDKPQEVYVQPDVRPQPPSP | 1167 |
| Qy | 1153 | REGPLPAARAGATLERAKTSLSPKGVVXVDVAFGGAVENPEYLTTCGGGAAPQHPPPA | 1212 |
| Db | 1168 | ENPFVSR-----KNGDLQ-----ALDNPEYHSSAG-----PPKA | 1198 |
| Qy | 1213 | -----FSPAFLNLYYWDQDPPPERGA--PPST | 1236 |
| Db | 1199 | EDEVNEPLYNTFTNALGNAEYKMSLLSVPEKAKKAFNDPDYWNHSLPRSTRLOHPDY | 1258 |
| Qy | 1237 | FKGTPT-----AENPEYL 1249 | |
| Db | 1259 | LOEYSTKYFYKQNGRIRPIVAENPEYL 1285 | |

DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (BC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthopterygii; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Wittbrodt J., Adam D., Malitschek B., Maueker W., Raulf F.,
 RA Telling A., Robertson S.M., Scharlt M.;
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RL inducing Tu locus in Xiphophorus.";
 RL Nature 341:415-421(1989).
 RN [2]
 RP REVISION TO 515.
 RA Scharlt M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright.. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X16891; CAA34770.2; -;
 DR PIR; S06142; S06142.
 DR HSSP; P11362; 1FGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
 FT KINASE.
 FT DOMAIN 26 642
 FT TRANSMEM 643 665
 FT POTENTIAL.
 FT DOMAIN 666 1167
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 710 977
 FT NP_BIND 716 724
 FT BINDING 724 743
 FT BINDING 743 743
 FT ACT_SITE 835
 FT BINDING 835 835
 FT DISULFID 195 204
 FT DISULFID 199 212
 FT DISULFID 220 228
 FT DISULFID 224 226

FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 256 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 311 BY SIMILARITY.
 FT DISULFID 315 330 BY SIMILARITY.
 FT DISULFID 333 337 BY SIMILARITY.
 FT DISULFID 504 513 BY SIMILARITY.
 FT DISULFID 508 521 BY SIMILARITY.
 FT DISULFID 524 533 BY SIMILARITY.
 FT DISULFID 537 553 BY SIMILARITY.
 FT DISULFID 556 569 BY SIMILARITY.
 FT DISULFID 560 577 BY SIMILARITY.
 FT DISULFID 593 615 BY SIMILARITY.
 FT DISULFID 618 626 BY SIMILARITY.
 FT DISULFID 622 634 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;
 Query Match 39.9%; Score 2715.5; DB 1; Length 1167;
 Best Local Similarity 45.4%; Pred. No. 1.7e-138;
 Matches 577; Conservative 162; Mismatches 387; Indels 145; Gaps 27;
 QY 4 AALCRWGLLLALLPPGAAST---OVCTGTOMKRLRASPETHLDMLRHLHYOGCOVQCN 59
 DB 8 AALQ--LLLVLSTSRCCSTDPDRKVCQGTNSQMTM---LDNHYLKKMKYSGCNVLEN 62
 QY 60 LELTYLPTNASLSLQDIQEVGYVLIAHNVQVPLQRLRIVRGTRQTFEDNYALAVLN 119
 DB 63 LEITYTQENQDLSFLQSIQEVGYVLIAMNEVSTIPLNLRILRQNLVEGNTLLVMSN 122
 QY 120 GDPLNNQYKANSK----FIGITELQRLSLTEILKGGVLIQRLNPOLCYODTILMKDIFHK 175
 DB 123 -----YQKNPSPDVYQVGLKQLQLSNLTEILSGGVKVSHPNLLCNVETINWMDIVDK 175
 QY 176 NNQLALTLDITNRSRACHPCSPCKSGRCWSESSEDCOSLRTVTCAGG--ARCKPLPTD 234
 DB 176 TSNPTMNLIPHAFERQCKDHGCVNGSCWAPGPGHCORFTKLLCAECCNRRCRGPKPID 235
 QY 235 CCEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVITYNTDTFESMPNPEGRYTFGAS 294
 DB 236 CCNEHCAGGCTGPRATDCLACRDFNDGCTCKDTCPPPKIYDIVSHQVVDNPNIKYTFGAA 295
 QY 295 CVTACPNYLSLDVSGSCTLVCPLNHNOEVTADGTQRCCKSKPCARVCYGLGMEHLREVR 354
 DB 296 CVKPCPSNVVYTE--GACVRSAGMLEVD--ENGKRSCKPCDGVCPKCDGIGIGLSNTI 353
 QY 355 AVTSANTIQEPAGCKKIFGLSFLPESFDGDPASNTAPLQPEQLQVFTLEETITGVLYISA 414
 DB 354 AVNSTNRSFSNCTKINGDIIILNRSFEGDPPHYKIGTMDPEHLMNLTVTKEITGVLYIMW 413
 QY 415 WPSLPLDSVFNQLQVIRGRILHNHAYS--LTLQGLIGISWLGRLSRLBELSGSLALHNH 473
 DB 414 WPNNTSLSVFQNLLEIIRGRTTFSRFGSVVVQVRLHQLMGLRSLKEYSAGNVILKNTLQ 473
 QY 474 LCFVHTVPWQLFRNPQALLHTANRPEDECVGGLACHQLCARGHCNGPGPTQCVNCQSQ 533
 DB 474 LRYANTINWRRLFRSEDSQIEYDART-----ENQTCNNECEDGCGNGPGPTMVCVSLH 526
 QY 534 FLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEPEADQCVACHYKDP 593
 DB 527 VDRGRCVASCNLLQGGPREAQVDGRCVQCHOELVQTDLSLTCTGPGPANCSKAHFDQD 586
 QY 594 PFCVARCPGKVPDLSYMPIWKPFDEEGACQPCPINCTHSCVDSCLDDKGCPEARASPLTS 653


```
Db 587 POCIPRPHGILGDGTL-IWKYADKMGQCQPCQHCNCTGCGSPGLSGCRGD-IVSHSSL 644
Qy 654 IVSAVVGILLVVLGVVFGILIKRRQOKIRKTYMRRLQETELVEPLTPSGAMPNQAMR 713
Db 645 AVGLVSGLLITVIALVLLVRRRIK-RRTIRCLLQELVEPLTPSGQAPNQAFLR 703
Qy 714 ILKETLRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYV 773
Db 704 ILKETEFKDRVLGSGAFGVYKGLWNPDCENIRIPVAIKVLRNTPSPKANKEILDEAYV 763
Qy 774 MAGVGSPYVSRLLIGLICLTSTVLTQMPYGCCLLDHVRNRLGSDLLNWCWQIAKGM 833
Db 764 MASVDHPHVCRLGICLTSAVLVTQMPYGCCLLDYVQHQERICGWLNLNWCWQIAKGM 823
Qy 834 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMWALE 893
Db 824 NYLEERHLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMWALE 883
Qy 894 ILRRRTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
Db 884 ILQWTVTHQSDVMSYGVYVWELMTFGSKPYDGIIPAKEIASVLENGERLPQPICTIEVYM 943
Qy 954 IMVKCMWIDSECPRELVESESRMARDQRFVWIONEDLGPASPLDSTFYRSLLEDD 1013
Db 944 IILKCMWIDSECPRELVESESRMARDQRFVWIONEDLGPASPLDSTFYRSLLEDD 1000
Qy 1014 MGDVDAEYLVPOQGFCCPDAPAGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPS 1073
Db 1001 --DVDAEYLVYKRI-----NRQS-----E 1021
Qy 1074 PLAPSEGAGSDVFDGLMGGAAGKLGSLTHDPSPLQRYSEDPTV-PLPSETDGYVAPLT 1132
Db 1022 PCIPPTGH-----PVRENSITLRNLSDTQNALEKLDGH----- 1056
Qy 1133 CSQPEYVNPQDVRPQ-----PSPRE-----GPLP-AARPAGATLERAKTLPSCQNG 1179
Db 1057 -----EYVNGPGETSRSLSDIYNPNYEDLTDGMPVSLSSQEAFTNFSREYLTNQNS 1111
Qy 1180 VVKDVEAFGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPRGAPPSTFKG 1239
Db 1112 L-----PLVSSGSMDDPY---QAG-----YQAFA-----LPQTGALTNGMF 1146
Qy 1240 TPTAENPEYLG 1250
Db 1147 LPAENLEYLG 1157

RESULT 9
ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (BC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90083234; PubMed=2687875;
RX Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RL overexpression in a subset of human mammary tumors.";
RN Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
```

```
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
FT POTENTIAL
FT RECEPTOR_PROTEIN_TYROSINE_KINASE_ERBB-3.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT PROTEIN_KINASE.
FT ATP (BY SIMILARITY).
FT ATP (BY SIMILARITY).
FT BY SIMILARITY.
FT BY SIMILARITY.
```

| | | | | |
|----|----------|---|------|---|
| FT | DISULFID | 210 | 218 | BY SIMILARITY. |
| FT | DISULFID | 214 | 226 | BY SIMILARITY. |
| FT | DISULFID | 227 | 235 | BY SIMILARITY. |
| FT | DISULFID | 231 | 243 | BY SIMILARITY. |
| FT | DISULFID | 246 | 255 | BY SIMILARITY. |
| FT | DISULFID | 259 | 266 | BY SIMILARITY. |
| FT | DISULFID | 280 | 301 | BY SIMILARITY. |
| FT | DISULFID | 305 | 320 | BY SIMILARITY. |
| FT | DISULFID | 323 | 327 | BY SIMILARITY. |
| FT | DISULFID | 500 | 509 | BY SIMILARITY. |
| FT | DISULFID | 504 | 517 | BY SIMILARITY. |
| FT | DISULFID | 520 | 529 | BY SIMILARITY. |
| FT | DISULFID | 533 | 549 | BY SIMILARITY. |
| FT | DISULFID | 552 | 565 | BY SIMILARITY. |
| FT | DISULFID | 556 | 573 | BY SIMILARITY. |
| FT | DISULFID | 576 | 585 | BY SIMILARITY. |
| FT | DISULFID | 589 | 610 | BY SIMILARITY. |
| FT | DISULFID | 613 | 621 | BY SIMILARITY. |
| FT | DISULFID | 617 | 629 | BY SIMILARITY. |
| FT | CARBOHYD | 126 | 126 | N-LINKED (GLUCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 250 | 250 | N-LINKED (GLUCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 353 | 353 | N-LINKED (GLUCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 408 | 408 | N-LINKED (GLUCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 414 | 414 | N-LINKED (GLUCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 437 | 437 | N-LINKED (GLUCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 469 | 469 | N-LINKED (GLUCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 522 | 522 | N-LINKED (GLUCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 566 | 566 | N-LINKED (GLUCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 616 | 616 | N-LINKED (GLUCNAC. .) (POTENTIAL). |
| FT | VARSPLIC | 141 | 183 | ETLSGVYIEKNDKLCMDTIDMDRIDRDRAEIVVKNGR SC -> GQPPVPSGLTQPAQDWLDDDDPRLTLSSASK VPTLAAV (IN SHORT ISOFORM). |
| FT | VARSPLIC | 184 | 1342 | MISSING (IN SHORT ISOFORM). |
| FT | CONFLICT | 560 | 560 | E -> G (IN REF. 2). |
| FT | CONFLICT | 1064 | 1064 | E -> G (IN REF. 2). |
| SQ | SEQUENCE | 1342 AA; 148097 MM; 7201E7F66CA374BD CRC64; | | |

Query Match 35.7%; Score 2430.5; DB 1; Length 1342;
Best Local Similarity 40.5%; Pred. No. 4e-123;
Matches 531; Conservative 192; Mismatches 459; Indels 129; Gaps 32;

| | | | |
|----|-----|---|-----|
| Qy | 10 | GLLIALLPPGAA--STQVCTGDMKLRPLASPETHLMLRLHYOGCVQVQGNLEUTYLP | 67 |
| Db | 11 | GLLFLARGSEVGNQAVCPGTLNGLSVTGDENQYQTYLYKURCEVVMGNLEIVLTGH | 70 |
| Qy | 68 | NASLSFLQDIQEVQGYVLIHQNVRQVPLQRLIRVGTQLFEDNYALAVLNDGDLNNQY | 127 |
| Db | 71 | NADLSFLQWIREVTGYVLVAMNEFTLPLNLRVVRGTVVYDGKFAIFVMLN----- | 122 |
| Qy | 128 | IKANSKFIGITELQRLSLEILKGGVLTORNPOLCYQDTILWKDIFHKNNQALTLIDTN | 187 |
| Db | 123 | YNTNSSH-ALRQLRLTQLEILSGGVYIEKNDKLCMDTIDMDRIDRDRAEIVVKNDR | 178 |
| Qy | 188 | RSRACHPCSPMKSRMGESDCQSLTRTVACGGC-ARCKGPLPTCCHEQCAAGCTG | 246 |
| Db | 179 | NGRSCPPCHEVCKG-RCHGPGSEDCQTLTKICAPQCNHCFGPNPQCCHDEAGGCG | 237 |
| Qy | 247 | PKHSDCLAHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNYLST | 306 |
| Db | 238 | PODTCFACRHFNDSGACVPCQPLVYNKLTFTQLEPNHTKYQYGGVGVASCPHFV-V | 296 |
| Qy | 307 | DVGSCTLVPLHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLUREVRAVTSANIQEPAG | 366 |
| Db | 297 | DQTSVRCAPDPDKMEVD-KNGLKMPCEGGLCPKACEGTGSG--SRFQTVDSNIDGFWN | 353 |
| Qy | 367 | CKKIFGSLAFPEPSFDGDPASNTAPLOEQVPELLEEITGYLYISAWPDSLPLSVFQ | 426 |
| Db | 354 | CTKLGNLDFLITGLNGDPWHKIPALDPKLNVRFTREITGYLNIQSWPHMNFVS | 413 |
| Qy | 427 | NLQVIRGRLHNGAYS-LTLOGLGSIWGLRSLRGSLALIHHTHLCFVHTVPMQDL | 485 |
| Db | 414 | NLTITGRSLYNRGFSLLIMKNLVNTSLGPFSLKEISAGRIYISANRQLCYHHSINWTKV | 473 |

| | | | |
|----|------|--|------|
| Qy | 486 | FRNPQALLHTA-NRPDEECVGEGLACHOLCARGHCWGPPTQCVNCOSFLRQOECVVEC | 544 |
| Db | 474 | LRGPTTEERLDIKHNRPRRDCVAEGKVCDDLCSGGCGWGPQCLSCRYNVRGCVCTHC | 533 |
| Qy | 545 | RVLQGLPREYVNRHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFCVPCPSGV | 604 |
| Db | 534 | NFLNGEPREFAEAEFCFSCHPECQMEGTATCNGSGSDTCAQACFRDGHPCVSSPHGV | 593 |
| Qy | 605 | KPDLSYMPIWKFPDEBEGACQPCPINCTHSCVDLDDKGCFAEORA-----SPLTISVSAVVG | 660 |
| Db | 594 | LG--AKGPIYKYPDVQNECRPCHECTQCKGPELQDCLGQTLVLGKTLHTLTALTVIAG | 651 |
| Qy | 661 | ILLVVVLGVVFGILIKRROQKIR-KYTMRLLOETVELPEPLTPSGAMPNOAQRILKETE | 719 |
| Db | 652 | --LVIFMMLGGTFLYWRGRRIQNKRAMRYRGERGIESIEPLDPS-EKANKVLARIFKETE | 708 |
| Qy | 720 | LRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVMAGVGS | 779 |
| Db | 709 | LRKLKVLGSGVFTVHKGVWIPGESSIKIPVICIKVIEDKSGQSFQAVTDHMLAIGLDH | 768 |
| Qy | 780 | PYVSRLLGICLTSTVQLVTQLMPYGLDHDVRENRLGSGQDLLNMCQIAKGMYSLEVD | 839 |
| Db | 769 | AHIVELLGLCGSSILQVLYPLGSLLDHVRQHRGALGPQLLLNWGVQIAKGMYLEEH | 828 |
| Qy | 840 | RLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRFR | 899 |
| Db | 829 | GMVHRNLAARNVLLKSPQVQVADFGVADLLPDDKQLLYSEAKTPIKMALESIHFGKY | 888 |
| Qy | 900 | THQSDVWYGVVWELMTFGAPYDGI PAREIPDILLEKGERLPPOPICTIDVYMWKWC | 959 |
| Db | 889 | THQSDVWYGVVWELMTFGAPYAGRLAEVDPDLLEKGERLAQPOICTIDVYMWKWC | 948 |
| Qy | 960 | MIDSECRFRFELVSEFSRMARDPQRFVVIQNEIDLGA---SPLDSTFYRSLLEDDMDG | 1016 |
| Db | 949 | MIDENIRTFTELANEFTRMARDPPRYLVIKRES-OPGIAPGEPHGLTNKKLEVELEP | 1007 |
| Qy | 1017 | LVDAEYLVPOQGFPCPDPAAGAGMVHRRHRSSTRSGGLDITLGLP-SEEAAPRSPL | 1075 |
| Db | 1008 | ELDLDLDEABED-----NLATTLGSLSLPVGTNLNRPGSQSL 1048 | |
| Qy | 1076 | APSEGAGSDVDGDLGMAAKGLQSLPTHD-PSPLQRYSEDTPVLP-----SETDGYV | 1128 |
| Db | 1049 | SPSSGY-MPMNQNLIGESCQESAVSGSERCPRVSLH-----PMRGLCLASESEGHV | 1101 |
| Qy | 1129 | A-----PLTCSPOE---VYNQPDVRPPSPREGP-----L 1157 | |
| Db | 1102 | TGSEAELOEKVSMCRSRSRSPRGRGSAHYSQRHSLTPTVPLSPGLEEDVNGYVM | 1161 |
| Qy | 1158 | PAARPAGATLERAKTLSP-GKNGVV-----KDVFAFGGAVENPEYLTPOGGAAPQHPH | 1210 |
| Db | 1162 | PDTHLKGTPSSREGTLSSVGLSSVLGTEEBED-----BEYEMNRRRRHSP-PHPH | 1212 |
| Qy | 1211 | PAFSPAFDNLVYWD-----QPPERGAPPTFKGTPTAENPEYL 1249 | |
| Db | 1213 | RPSSLEELGYEYMDVGSGLSASLQSCPLHPVIMPTAGTTPDEDYEYM 1263 | |

RESULT 10
ERB3 RAT
ID ERB3 RAT STANDARD; PRT: 1339 AA.
AC Q62799; Q62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN NCB1_TaxID=10116;
RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
recombinant protein";
RL Gene 165:279-284 (1995).
RN [2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuroligins and their putative receptors, ErbB2 and
ErbB3, is induced during Wallerian degeneration";
RL J. Neurosci. 17:1642-1659 (1997).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTA.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; U29339; AAC28498.2; -;
DR EMBL; U52530; AAC53050.1; -;
DR HSSP; P11362; LFGLK
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; pkinase-like; 1.
DR Pfam; PF01030; Recept_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339
FT DOMAIN 20 643
FT TRANSMEM 644 662
FT DOMAIN 663 1339
FT DOMAIN 183 259
FT DOMAIN 707 964
FT NP_BIND 713 721
FT BINDING 740 740
FT ACT_SITE 832 832
FT DISULFID 186 194
FT DISULFID 190 202
FT DISULFID 210 218
FT DISULFID 214 226
FT DISULFID 227 235

FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1028 1028 L -> P (IN REF. 3).
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5P2402BDFD1E CRC64;

Query Match 34.7%; Score 2359.5; DB 1; Length 1339;
Best Local Similarity 40.8%; Pred. No. 2.6e-119;
Matches 523; Conservative 172; Mismatches 433; Indels 155; Gaps 34;

QY 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMLKRLPASPTHLDMLRHLRYQGCQVQGN 59
DB 7 LQVLC----FLLSLARGSEMGNSQAVCPGTGLNGSLVTGDNADNQYQTLKYKECEVWGN 62
QY 60 LEITYLPTNASLFLQDIQEVCYVLIHNNQVQPLQRLRIVRGTLQFEDNYALAVLDN 119
DB 63 LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSVLPLNLRVVRGTVQYDGKFAIFVMLN 122
QY 120 GDPLNNQYIKANSKFGITELQRLSITELKGGVLTORNPOLCYQDTILWKDIFHKNNQL 179
DB 123 -----YNTNSSH-ALRQLKFTQLTEILSGGVYIEKNDKLCMDTIDRVRVR--- 170
QY 180 ALTLIDTNRSRACHPCSPCKSGRCGSESSDCQSLTRTYVCAGC-ARCKGPLPTDCCHE 238
DB 171 GAEIVVKNNGANCPCHVECKG-RCMGPGDDCQIILTKICAPQCNGRCFGPNQCCHD 229
QY 239 QCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTTFESMPNDEGRYTFGASCVTA 298
DB 230 ECAGGCGSPQDTCFACRRFNDSGACVPRCPFLVYNKLTFOLEPNPHTKYQYGGVCVAS 289
QY 299 CPYNYLSTDVGSCTLVCPHLNQVTAEDGTQRCCKSKPCARVCYGLGMHLEHREAVTS 358
DB 290 CPNFFV-VQDTFCVRACPPDKMEVD-KHGLKMCPCGGLCPKACEGTGSG--SRYQTVD 345
QY 359 ANIQEFAGCKKIFGSLAFIPESPDGPASNTAPLQPEQLQVFTLEITCYLVISAMPDS 418
DB 346 SNIDGFVNCTKILGNLDFLITGLNVDPWHKI PALDPEKLVNFTREITGLINQSWPPH 405
QY 419 LPDLSVFQNLQVIRGRILHNGAYS-LTLOGLGISWLGRSLRSLRGSLALIHHTHLCFV 477
DB 406 MHNFSVFSNLTTIGRSLYNRGFSLLIMKNLNVTSIGFRSLKEISAGRVVISANQQLCYH 465
QY 478 HTVPWDLPRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGTQCNCQFLR 536
DB 466 HSLNMTLLRLLRGFSEERLDIKYDRPLGECIAEGRVKVCDPLCSSGGCGWGPQGCISCRNYSR 525
QY 537 GQECVEECRVLOGLPREYINARHCLPCHPECPONGSVTCFGEADQACVACAHYKDPFFC 596
DB 526 EGVCVTHCNFLOQEPREFVHEAQCFSCHEPCLPMEGTSTCNGSGGSACARCAUFRDOPHC 585
QY 597 VARCPGSKVPDLISYMPINWKPDEEGACQPCPINCTHSC--VDLDDKGCFAEQRASPLTSI 654

```
Db 586 VNSCPHGILG--AKGPIYKPAQNECRPCHENCQTQCGNPGELQCLGAEVLMKPHLV 643
Qy 655 VSAAVGIILVVVLGVVFGILIKRQOKIR-KYTMRRLLQETELVPLTPSGAMPQAQR 713
Db 644 IAVTVG--LAVILMILGGSFLYWRGRIQNKRAMRYLGRGSIPLDPS-EKANKVLAR 700
Qy 714 ILKETELRKVKVLGSGAGTGYVKGWIWDGENVKIPVAIKVURENTSPKANKEIIDEAV 773
Db 701 IFKETELRKVLGSGVGTGVHKGWIWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHMLA 760
Qy 774 MAGVGSPPVRLGLCLTSTVOLVTQMLPYGCLLDHVNRGRGLSGQDLNMCWQIAKGM 833
Db 761 VCSLDHAHIVRLGLGCPSSQLQVLYPLGSLLDHVHQHRTLPQLLLNMGVQIAKGM 820
Qy 834 SYLEDVRLVHRDLAARNVLKSPNPKVITDFGLARLLDIDETEHADGKGVPIKMALES 893
Db 821 YYLEEHSVHRDLALRNVLKSPSQVQVADFVADLLPPDDKQLLHSEAKTPIKMALES 880
Qy 894 ILRRFTHQSDVMSVGVTVWELMTGCAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYM 953
Db 881 IHFGKYTHQSDVMSVGVTVWELMTGCAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYM 940
Qy 954 IMVKCWMIDSECRPRFRELVSFMRMDPQRFVVIQNEIDLGPASPLDSTFYRSLEDD 1013
Db 941 VMVKCWMIDENIRPTFKELANEFTMRADPPRYLVIKRAS-GPGTP--PAAPSPVLTTK 997
Qy 1014 MGDVDAEYLVPOQGFPCPDPAAGAGGMVHRHRSSTRSGGDLTLGLEPSE- 1068
Db 998 L-----QEALEPEL-----DLDLDEAEESGLATS 1023
Qy 1069 -----EAPRSLAPSEG-----AGSDVFDGLMGAAKGLQSLPHD 1105
Db 1024 LGSALSPTGLTRPRGQSLLSPSSGPMNQSSGLGACLDLSAVLGGREGQFSRPSISLH- 1082
Qy 1106 PSPLQRYSEDPTPLPSTDGVV-----APL-----TC-----SPQPE-----VYNQPDV 1145
Db 1083 PIPGR-----PASESEGHVTGSEAELEKQVSCRSRSPRGRGSAVHSQRHS 1135
Qy 1146 RPQPPSPREGP-----LPAARPAGATLERAKTLSP-KGNQGVV-----KDVPAF 1187
Db 1136 LLTPVTPLSPGLEEDGNGVMPDTHLARGASSREGTLSSVGLSVLTGTEED- 1191
Qy 1188 GGAVENPEYLTPOGGAAPQHP 1210
Db 1192 ----EYEMRMRKRGSP-PRPP 1209

RESULT 11
EGFR_DROME
ID EGFR_DROME STANDARD; PRT: 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
RX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN (2)
RP REVISIONS.
```

```
RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN (4)
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; Tissue=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Glazer L., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN (5)
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN (6)
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yeaman M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaisi M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.S., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN (7)
RP SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=Daekwanryeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal
```



```

Db 609 ISNAYK--FDNRKTKICHECR-----TCNAGADHCQECVHVRDQGHVSECPKKNYN 660
Qy 602 -----SGVK-----PDLSPYPIW 614
Db 661 DRGVCRECHATCDGCTGPKDTIGIGACTTCNLAINDATVKRCLLKKDKCPD-GY--FW 717
Qy 615 KF--PDEGACQP-----CPI-----NCTH----- 632
Db 718 EYVHPQEOGSLKPLAGRAVCRKCHPLCELCTNYGYHEQVCKSKCTHYKRRQECETCPADH 777
Qy 633 -----SC-----VOLDDKG-----CPAQ 646
Db 778 YTBQRECFORHPBECNCTGPGADDCKSCRNFKFDANETGYPVNMSTFMFCTSKCPLEM 837
Qy 647 R-----ASPLTS-----IVSAVVGILLVVVLGVFGILIKRQ 679
Db 838 RHVNYQYTAIGPYCAASPRRSKITANLDVNMIFIITGAVLVPTICILCV--TYICRQK 895
Qy 680 QKIRKYT--MRRLLQETELVEPLTPSGAMPNOAQRILKTELKRVKVLGSGAGTVVYKG 737
Db 896 QKAKKETVMTMALSGCSDSEPLRPSNITGANLCKLIRIVKDAELRKGVLGMGAFGRVYKG 955
Qy 738 IWIPOGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVSPYVSRLLIGLICLTSTVOLV 797
Db 956 VWPVEGENVKIPVAIKELLKSTGAESSEFEFLREAYIMASEBHVNLKLLAVCMSSQMWLI 1015
Qy 798 TQMPYGCGLDVRNRRGLSGODLLNMCIOAKGMSYLEVDRLVHRDLAARNVLKSPN 857
Db 1016 TQMLPLGCLLDYVNRNRDKISGKALLNWSIOAKGMSYLEKRLVHRDLAARNVLVQTPS 1075
Qy 858 HVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRRFTHQSDVMSYGVTVWELMT 917
Db 1076 LVKITDFGLAKLLSSDSNEYKAAGSKMPIKWLALCIRNRVFTSKSDVWAFGVTIWEILL 1135
Qy 918 FGAKPYDGIPIARETIPDLLEKGERLPDPDICTIDVYIMVWCMIDSECRPRFRELVSFES 977
Db 1136 FGQPHENIPAKDIPDLIEVLGKLEQPEICSLDIYCTLSCWHLDDAAMRPTFKQLTVFA 1195
Qy 978 RMARDPORFVVIQNEGLG--PASPLDSTFYRSLLEDD---DMGDLVDAAEYLVPOQGFPC 1032
Db 1196 EFARDPGYLAIPGDKFRLPA-----YTSQDEKILIRKLAPTTDGEAIAKPDYLIQ 1248
Qy 1033 PDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEEAPEAP-----RSLPAPSEAGSDVVF 1086
Db 1249 PKAAPGPS-----HRTDCT-----DEMPKLNRYCKDPSNKNSSTGDDER 1287
Qy 1087 DG---DLGMAAGKGLQSLPHTDPSLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNP 1143
Db 1288 DSSAREVGVGNLR-----LDLPVDEDDYLMPT-TCQPGPNNNNM 1325
Qy 1144 DVPQPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVKDVFAGGAVENPEYL----TP 1199
Db 1326 N-----NPNQNNAAVGAAGYM-----DLIGVPVSDNPEYLLNAQTL 1364
Qy 1200 QGGAAPQH-----PPAFSP-AFDNLYYWD 1224
Db 1365 VGGESPIPTOTIGIPVMGGPGTMEVKVPMGPSEPTSSDHEYND 1408

RESULT 12
ERBB ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses: Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilleen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RB Tittenden L.B., Raines M.A., Kung H.-J.,
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFVLV.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE-TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
FT SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
Query Match 25.7%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 6e-87;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;
Qy 587 CAHYKDPFPCVARGSPGVKPDLSYMPIWKFPEEGACQPCPINCTHSCVDLDDKGPASQ 646
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VWKYADANAVCOLCHPNCNTRGCKGPGLEGCP--- 58
Qy 647 RASPLTSTVSAAV-GILLAVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSCA 705
Db 59 NSKTPSTAAGVGLCLVVGIGIGLYLRRR-HIVRRKTLRLLRQLRELVEPLTPSGE 117
Qy 706 MPNOAQRILKETELKRVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 765
Db 118 APNOAHLRLAKETEFKKVKVVLGSGAFGTVYKGLWIPEGEKVIPIVAIKELREATSPKANK 177
Qy 766 EILDEAYVMAGVSPYVSRLLIGLICLTSTVQLVLTQMPYGCGLDVRNRRGLSGODLLNM 825
Db 178 EILDEAYVMASVDNPHVCRLLIGLICLTSTVQLITQMPYGCGLDVRNRRGLSGODLLNM 237
Qy 826 CMOJAKGMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 885
Db 238 CVQJAKGMYLDEERRLVHRDLAARNVLKTPHVKITDFGLAKLGLGADEKEYHAGGKVP 297
Qy 886 IKMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIARETIPDLLEKGERLP 945

```

Db 298 IKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIASEISSVLEKGERLPQPP 357
QY 946 ICTIDVYIMVWKWMDSCRRPRERELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVWKWMDADSRKFRKELIAEFSKWARDPPRYLVIOQDERMHLPSPTDSKF 417
QY 1005 YRSLLDEDDMGDLVDAEYLVPOQGFCDPAPGAGGMVHRRHSRSTSGGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1065 PSEEEAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLFTHDPSPLOYSDEPTVP 1119
Db 450 -----SRTLPLLSLSATSNSATNCID-----RNGQGHVPVREDSFVQYSSDPTGN 495
QY 1120 LPSET--DGYVAPLTCSPQEVYVQPDVRPQPSREGPLPAARPAGATLERAKTLPCK 1177
Db 496 FLEESIDDGFL-----PAPEYVNO--LMPKXPS-----TAMVQ 526
QY 1178 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLXY 1222
Db 527 NQIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPY 578
QY 1223 WQO-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
Db 579 WIQSGNHQINLNDPQYQDPLPNETKPNGLLKVPAEAENPEYLRVAAP 625

RESULT 13
ERBB AVIER
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family";
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene";
RL Science 224:1456-1459(1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; --
DR PIR; A00644; TVYUH.
DR HSSP; P11362; LEKG.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBDD06745D609 CRC64;

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 1.8e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

QY 587 CAHYKDPPECVACPSGVKPDLSYMPWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
Db 3 CAHFDGPHCVKACPAVLGENDTL-VRYADANAVCOLCHPNCTRGCKPGLEGCP--- 58
QY 647 RASPLTIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGA 705
Db 59 NSGKTPSIAAGVVGGLLVVGLGIGLYLRR-HIVRKTRLLRLQERELVEPLTPSGE 117
QY 706 MPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANK 765
Db 118 APNQAHRLILKETEFKVKVVLGSGAGFTYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
QY 766 EILDEAYNAGVSGPVSVRLGICLTSTVOLTPMLPYGCLLDHVRNRRGLCSQDLLNW 825
Db 178 EILDEAYNAGVSDNPHVCRLLGICLTSTVOLTPMLPYGCLLDYIREHKNIGSQYLLNW 237
QY 826 CMQIAGKSYLEDVRLVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGCKVP 885
Db 238 CVQIAGKMYLBERLVRDLAARNVLVKTPOHVKITDFGLAKLLGADKEYHAEGCKVP 297
QY 886 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAEIPDLEKGERLPQPP 945
Db 298 IKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQPP 357
QY 946 ICTIDVYIMVWKWMDSCRRPRERELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVWKWMDADSRKFRKELIAEFSKWARDPPRYLVIOQDERMHLPSPTDSKF 417
QY 1005 YRSLLDEDDMGDLVDAEYLVPOQGFCDPAPGAGGMVHRRHSRSTSGGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1065 PSEEEAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLFTHDPSPLOYSDEPTVP 1119
Db 450 -----SRTLPLLSLSATSNSATNCID-----RNGQGHVPVREDSFVQYSSDPTGN 495
QY 1120 LPSET--DGYVAPLTCSPQEVYVQPDVRPQPSREGPLPAARPAGATLERAKTLPCK 1177
Db 496 FLEESIDDGFL-----PAPEYVNO--LMPKXPS----- 524
QY 1178 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPHPPAFSPAFD 1218
Db 525 --VQNIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFE 574
QY 1219 NIYYWQDPPPERGAPPSTFKGPTTAENPEY 1248
: : : : :
: : : : :
: : : : :

or send an email to license@isb-sib.ch).

```
CC -----
CC EMBL; M20386; AAA48760.1; ..
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN KINASE TYR; PARTIAL.
DR PROSITE; PS00111; PROTEIN KINASE DOM; PARTIAL.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 667 POTENTIAL.
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON TER 703 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Query Match 23.7%; Score 1614; DB 1; Length 703;
Best Local Similarity 44.3%; Pred. No. 1.3e-79;
Matches 313; Conservative 113; Mismatches 253; Indels 28; Gaps 11;

QY 8 RWGLLLALLPPGAA-----STOVCTGTDMKRLPASPETHLDMLRLHYQGVVQGNLE 61
DB 13 RGAALVLLLLGVALCSAVEEKVCCQGTNNKLTOLGHVEDHFTSLQRMNNCEVLSNLE 72

QY 62 LTYLPTNASISFLQDIOEVGYVLIHQNVRQVPLQRLIRVGTQLFEDNYALAVLDNGD 121
DB 73 ITVVEHNRDLTFLKTIQEVAGYVLIHNMVDVPLENLQIRGNVLYDNSFALAVLSNYH 132

QY 122 PLNNQYIKANSKFIGITELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKNNQAL 181
DB 133 MNKTQ-----GLRELPMKRLSEILNGGVKISNNPKLNCNMDTVLWNDIIDTSRK-PL 182

QY 182 TLID-TNRSRACHPCSPMKGSCWGESSEDCQSLTRTVCGAGCA-RCKGPLPTDCCHEQ 239
DB 182 TLID-TNRSRACHPCSPMKGSCWGESSEDCQSLTRTVCGAGCA-RCKGPLPTDCCHEQ 239
```

```
DB 183 TVLDPASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQOCSGRCGRKVPSPDCCHQ 242
QY 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGYTTFGASCVTAC 299
DB 243 CAAGCTGPRSDCLACRKRFRDDATCKTCTPLVLYNPTTYOMDVNPEKYSFGATCVREC 302
QY 300 PYNLYSTDVGSCTLVCPLHNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTS 359
DB 303 PHNYVTVDHSGSVRSCTNDTVEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGLSINAT 361
QY 360 NIQEFACCKIFGSLAFIPESFDGDPASNTAPLOPEQLOVFEETLEEITGYLYISAWPDSL 419
DB 362 NIDSFKNCTKINGDVLSILPVAFLGDAFTKTLPLDPKLDVFTVKESISGFLLOAWPDNA 421
QY 420 PDLVSFONLOVIRGRILHNHNGAYSUTLQGLGTSWLGRLSLRELGLSGLALIHNTLHLCFVHT 479
DB 422 TDLYAFENLEIIRGTRTKHQGOYSLAVNVLKIQSLGLSLKEISGDITAIMKNKULCYADT 481
QY 480 VPMQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCGWGPPTQCVNCSQFLROQE 539
DB 482 MNWRSLEFATOSQTKIIQNRKNKNDCTADRHVCDPLCSDVGCWGPDPHCFSCRFPSRQKE 541
QY 540 CVBECEVLQGLPREYVNAHCLPCHPECPQNG---SVTCFGEADQCVACAHYKDPFPC 596
DB 542 CVKQCNLTQGEPRERFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCKMCAHFIDGPHC 601
QY 597 VARCPGKVPDLSYMPWIKFPDEBACQPCPINCTHSCVDLDDKCPAEQASPLTSIVS 656
DB 602 VKACPAGVLGENDTL-VKKYADANAVCOLCHPNCTRGCKGPGLEGCP---NGSKTPSIAA 657
QY 657 AVV-GILLVVVLGVVFGILIKRROOKIRKYTMRLRLOETELVEPLTP 702
DB 658 GWVGGLLCLVVVGLIGLGLYLRRL-HIVRKRTLRLLOERELVEPLTP 703
```

Search completed: July 22, 2003, 08:44:24
Job time : 20.2304 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-103-117-12

Perfect score: 6809

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:*
1: sp_arChaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 6253 | 91.8 | 1259 | 6 O18735 | O18735 canis famil |
| 2 | 3156 | 46.4 | 1209 | 11 Q9QX70 | Q9QX70 rattus norv |
| 3 | 3128 | 45.9 | 1210 | 11 Q9EP98 | Q9EP98 mus musculu |
| 4 | 2749 | 40.4 | 1165 | 13 Q9YH40 | Q9YH40 xiphophorus |
| 5 | 2729.5 | 40.1 | 1137 | 13 Q9W6F6 | Q9W6F6 gallus gall |
| 6 | 2315 | 34.0 | 1328 | 13 P79754 | P79754 fugu rubrip |
| 7 | 2047.5 | 30.1 | 1433 | 5 Q8I8H9 | Q8I8H9 anopheles g |
| 8 | 1798.5 | 26.4 | 419 | 4 Q9UK79 | Q9UK79 homo sapien |
| 9 | 1739 | 25.5 | 367 | 11 Q8R2X1 | Q8R2X1 mus musculu |
| 10 | 1720 | 25.3 | 729 | 15 Q86712 | Q86712 avian rous- |
| 11 | 1718 | 25.2 | 567 | 15 Q86714 | Q86714 avian rous- |
| 12 | 1697.5 | 24.9 | 412 | 4 Q8WYV0 | Q8WYV0 homo sapien |
| 13 | 1653.5 | 24.3 | 962 | 15 Q64895 | Q64895 avian eryth |
| 14 | 1645 | 24.2 | 545 | 15 Q85468 | Q85468 avian eryth |
| 15 | 1519.5 | 22.3 | 655 | 11 Q9WVF5 | Q9WVF5 mus musculu |
| 16 | 1503.5 | 22.1 | 643 | 11 Q9ERV6 | Q9ERV6 mus musculu |

17 1288 18.9 1193 5 Q9Y1X8
18 1215.5 17.9 1368 5 Q23821
19 1165 17.1 1717 5 Q26566
20 1145 16.8 527 13 Q90836
21 1014.5 14.9 478 11 Q9ESE0
22 961.5 14.1 599 13 Q9PSH2
23 906 13.3 165 4 Q14256
24 887 13.0 176 11 Q923V5
25 806.5 11.8 346 13 F11776
26 778 11.4 435 5 Q8SZW1
27 758.5 11.1 1362 13 Q9PV24
28 754.5 11.1 311 13 Q9N162
29 729 10.7 1671 5 Q9NUV5
30 724 10.6 331 4 Q9BUD7
31 723 10.6 149 6 Q9BG66
32 703.5 10.3 1368 13 Q8UW85
33 697 10.2 1418 13 Q93457
34 683.5 10.0 1369 13 Q8UW86
35 674 9.9 1358 13 Q73798
36 669.5 9.8 1472 5 Q9U5A8
37 657 9.6 1412 13 Q8UW84
38 643.5 9.5 1418 13 Q8UW83
39 640 9.4 1245 13 Q9YGH8
40 640 9.4 1371 11 Q9QVW4
41 621.5 9.1 2144 5 Q9VD94
42 598 8.8 987 11 Q91YM0
43 595 8.7 935 4 Q96L35
44 595 8.7 987 11 Q99MR2
45 587.5 8.6 1036 4 Q07912

ALIGNMENTS

RESULT 1

O18735
ID O18735 PRELIMINARY; PRT: 1259 AA.
AC O18735;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ErBb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbB-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; Ylp_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Q9Y1X8 ephydatia f
Q23821 caenorhabdi
Q26566 schistosoma
Q90836 gallus gall
Q9ese0 rattus norv
Q9psH2 gallus gall
Q14256 homo sapien
Q923V5 rattus norv
P11776 xiphophorus
Q8szw1 drosophila
Q9pv24 xenopus lae
Q9162 xiphophorus
Q9njv5 biomphalari
Q9bud7 homo sapien
Q9bg66 oryctolagus
Q8uw85 paralichthy
Q93457 scophthalmu
Q8uw86 paralichthy
Q73798 xenopus lae
Q9u5a8 bombyx mori
Q8uw84 paralichthy
Q8uw83 paralichthy
Q9ygh8 scophthalmu
Q9qvw4 rattus sp.
Q9vd94 drosophila
Q91ym0 mus musculu
Q96l35 homo sapien
Q99mr2 mus musculu
Q07912 homo sapien

```
Query Match      91.8%; Score 6253; DB 6; Length 1259;
Best Local Similarity 91.6%; Pred. No. 0;
Matches 1154; Conservative 40; Mismatches 60; Indels 6; Gaps 2;

QY 1 MELAALCRWGLLLALLPQCAASTOVCTGTDMLKRLPASPTHLDMLRHLYQCCVVQGNL 60
DB 1 MELAALCRWGLLLALLPQCAAGTQCTGTDMLKRLPASPTHLDMLRHLYQCCVVQGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQSVLIHQAQVPLQRLIRVRGTQLFEDNEDYALAVLNG 120
DB 61 ELTYLPANASLFLQDIOEVQSVLIHQAQVPLQRLIRVRGTQLFEDNEDYALAVLNG 120
QY 121 DPLNNQYIKANSKFTIGTELQRLSLEILKGVLIQRLPOLCYQDTILWKDIFHQNQLA 180
DB 121 DPLEGGIPAPGAQOGLRELQRLSLEILKGVLIQRLPOLCYQDTILWKDIFHQNQLA 180
QY 181 LTLIDNRSRACHPCSPMCKSGRCSESDCQSLTRTVACGCCARCKGPLPTDCCHEQC 240
DB 181 LTLIDNRSRACHPCSPMCKSGRCSESDCQSLTRTVACGCCARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP 300
QY 301 YNLSLTDVGSCTLVCPLNQOEVTAEQDTCQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
DB 301 YNLSLTDVGSCTLVCPLNQOEVTAEQDTCQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFELEITGVLYISAWPDSL 420
DB 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFELEITGVLYISAWPDSL 420
QY 421 DLSVFQNIQVIRILHNGAYSLTIOGLIGISWGLRSRLGSLALIHNTLHCFVHTV 480
DB 421 NLSVFQNIQVIRILHNGAYSLTIOGLIGISWGLRSRLGSLALIHNTLHCFVHTV 480
QY 481 PWDQLFRPHQALLHTANRDECEVGEGLACHOLCARGHCWGPQTQVCNCSQFLRGQEC 540
DB 481 PWDQLFRPHQALLHTANRDECEVGEGLACHOLCARGHCWGPQTQVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVYNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
DB 540 VEECRVLQGLPREYVYNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
QY 601 PSQVVKPDLISYMPIWKPFDEEGACQPCINCTHSCVDLDDKGPAPQASPLTISVAVVG 660
DB 600 PSQVVKPDLISYMPIWKPFDEEGACQPCINCTHSCVDLDDKGPAPQASPLTISVAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
DB 660 ILLVVVLGVVFGILIKRQOKIRKYTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
DB 720 RKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGICLSTVOLVTQMPYGCCLLDHVHRENRGLSGQDLNMCQIAKMSYLEDR 840
DB 780 YVSRLLGICLSTVOLVTQMPYGCCLLDHVHRENRGLSGQDLNMCQIAKMSYLEDR 840
QY 841 LVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKVPDKWMALESILRRFT 900
DB 840 LVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKVPDKWMALESILRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
DB 900 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
QY 961 IDSECRPRFRELVEFSFMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
DB 960 IDSECRPRFRELVEFSFMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1019
QY 1021 EBYLVPOQGFPCDPAACAGGMVHRRSSSTRSGGDLTLGLFSEEPKSLAPSEG 1080
DB 1020 EBYLVPOQGFPCPEPTGAGGTFAHRRSSSTRSGGDLTLGLFSEEPKSLAPSEG 1079
QY 1081 AGSDVFDGDLGMAAGKQLSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
DB 1080 AGSDVFDGDLGMAAGKQLSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1139
QY 1141 NQDVRPQPPSPREGPLPAARPAAGATLER-----AKTILSPGKNGVYKDVFAFGGAVENPE 1195
DB 1140 NQPEWQPPPLALEGLPPSRPACATLERPKTILSPKTLSPGKNGVYKDVFAFGGAVENPE 1199
QY 1196 YLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPSPFTKGTPTAENPEYGLDVPV 1255
DB 1200 YLAPRGRAAPQHPHPPAFSPAFDNLVYWDQDPPSERGSPFTPEGTPTAENPEYGLDVPV 1259

RESULT 2
QYQX70 PRELIMINARY; PRT; 1209 AA.
AC Q9QX70;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Epidermal growth factor receptor.
GN EGFR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RX MEDLINE=9025888; PubMed=2342466;
RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA Earp H.S.;
RT "A truncated, secreted form of the epidermal growth factor receptor is
RT encoded by an alternatively spliced transcript in normal rat tissue.";
RL Mol. Cell. Biol. 10:2973-2982(1990).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Petch L.A.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Guttridge K., Dawson T.L., Earp H.S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; M37394; AAF14008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1209 AA; 134891 MW; 96FEE7F6CCLB7773 CRC64;

Query Match      46.4%; Score 3156; DB 11; Length 1209;
Best Local Similarity 50.0%; Pred. No. 3.2e-229;
Matches 639; Conservative 168; Mismatches 356; Indels 114; Gaps 26;

QY 3 LAALCRWGLLLALLPQCAASTOVCTGTDMLKRLPASPTHLDMLRHLYQCCVVQGNLE 61
DB 3 LAALCRWGLLLALLPQCAAGTQCTGTDMLKRLPASPTHLDMLRHLYQCCVVQGNLE 61
```


KW ATP-binding; Receptor; Transferase.

SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 45.9%; Score 3128; DB 11; Length 1210;

Best Local Similarity 49.5%; Pred. No. 4.2e-227;

Matches 630; Conservative 170; Mismatches 362; Indels 110; Gaps 23;

```
QY 11 LLLALLPPGAA--STQVCTGTOMKRLRASPETHDMLRHLYGCGCQVOVQGNLELTLYPTN 68
DB 14 LLTALCAAGALEEKVKCGTNRLTQLGTFEDHFLSLQRMVYNNCEVLGNLEITYVQRN 73
QY 69 ASLSFLDDIQEOGVYVLIHNVQVPLQRLRIVRGTQTFEDYALAVLDNGDPLNNQVI 128
DB 74 YDLSFLKTIQEVAGYVLIALTVERIPLENLQIRGNALYENTYALAILSN----- 124
QY 129 KANSKFIGITELQRLSLTEILGGVLIQRNPOLCYODTILWKDI-----FHKNQLALTLI 184
DB 125 -YGTNRGTGURELPMRLNQLIELIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMMDL--- 180
QY 185 DTRSRACHPCSPMCKGRSCWGESSEDCQSLTRTVCAAGCA-RCKGPLPTDCCHEQCAAG 243
DB 181 -QSHPSCKPKCPSCPNCGWCGGGEENCQKTKIICAQCSHRCGRSPSDCCNCAAG 239
QY 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDIFESMPNPEGRVTFQASCVTACPNY 303
DB 240 CTGPRESDCLVCQKFQDEATCKDTCPLMLNPTTYQMDVNPPEGKYSFGATCVKCKPRN 299
QY 304 LSTDVCSCTLVCPHLNQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRATVSANIOE 363
DB 300 VYTDHGSCVRACGPDYEV-EEGIRKCKKCDGPKRCVNGIGIGEFKDTLSINATNIKH 358
QY 364 FAGCKKIFGSLAFLPSFQDGPASNTAPLQPEOLQVFETLEEITGYLXISAWPDSLPDLS 423
DB 359 FKYCTAISGDLHLPLVAFKGSDFTRTPPLDPRELEIKTVKEITGELLQAMPDNDTDLH 418
QY 424 VFONLOVIRGRLHNGAYSLLTLOGLIGISWLGRLSRELGSGLALIHNNTHLCFHVTPWD 483
DB 419 AFENLEIRGRTKQHQFQSLAVVGLNITSLGRLSKEISDGDVIIISGRNLCVANTINWK 478
QY 484 QLFNRPHQALLHTANPEDECVEGEGLAGHOLCARGHMGPGPTQCVCNCSQFLRGQCVEE 543
DB 479 KLFGTNQTKTMNNAEADCKAVHVNCPNPLSCGCGWGPEDRDCVSCQNVSGRECVBK 538
QY 544 CRVQLGPREYVNAHCLPCHPECPQNGSVTFCGPEADQCVACAHYKDPFPFCVACRCPG 603
DB 539 CNILEGEPREFVENSEICQHPCLPOAMNITCTGRGPDNCIQAHYIDGPHCVKTCPCAG 598
QY 604 VXPDLSYMIFKFPDEEGACQPCPINCTHSCVDLDDKGPAPORASPLTSIVSAVVGILL 663
DB 599 IMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCCEVWPSPGPKIPSIATGIVGGL 657
QY 664 VVVLGVVFGI-LIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAOQMRILKETELRK 722
DB 658 FIVV-VALGIGLFMRRRHVRKTLRLRLQERELVEPLTPSGEAPNQAHURLIKETEFKK 716
QY 723 VKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPPKANKEILDEAYVWAGVSPYV 782
DB 717 IKVLSGAGFVYKGLWIPGEGKVPVPAIKELREATSPKANKEILDEAYVWASVDNPHV 776
QY 783 SRLLGICLTSTVOLVTQMPYGLLDHVRNCRGLSQDLLNWCQIAKGMYSYLEVRLV 842
DB 777 CRLLGICLTSTVOLITQMPYGLLDYVREHKDNIQSQYLLANVCVQIAKGMNLYEDRLV 836
QY 843 HRDLAARNVLKSPNPKVITDFGLARLLDIDETEHADGKGKVPKMALESILRRFTHQ 902
DB 837 HRDLAARNVLVITPQHVKITDFGLAKLLGAEEKYEHAEGKVPKMALESILHRIYTHQ 896
QY 903 SDVWSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMVVKCWMID 962
DB 897 SDVWSYGVTVWELMTGSKPYDGIPIASDISILEKGERLPQPPICTIDVYIMVVKCWMID 956
QY 963 SECRPRFRELVEFSFMSARDPQRFVVIQ-NEDLGASPLDSTFYRSLLEDMDGDLVDAAE 1021
DB 963 SECRPRFRELVEFSFMSARDPQRFVVIQ-NEDLGASPLDSTFYRSLLEDMDGDLVDAAE 1021
```

```
DB 957 ADSRPFKRELIILEFSKWARDPQRYLVIOGDERMHLPSPDTSNFYRALMDEEDMEDVVVDAD 1016
QY 1022 EYLVPQOQGFCDPAPGAGGMVHRRHSSSTRSGGDLTLGLFSPSEEAAPRSLAPSEGA 1081
DB 1017 EYLTPQOQGF-----NSPST-----SRTPLLSLSA 1042
QY 1082 GSDVDFDGDILGMAAGKQLSLTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEY 1139
DB 1043 TSN-----NSTVACINRNGSCRVKEDAFQRYSSDPTGAVTEDNIDDAFL-----PUPEY 1092
QY 1140 VNQPDVRPOPSPREGPLPAARPAGATLERAKTLSGKNGVVKDVFAGFGAVENPEYL-T 1198
DB 1093 VNQ-SVPRKPPAGSVQNPVYHNPQLHP-----AFGRDLHYQN--PHSNAVGNPEYLN 1141
QY 1199 PGGAAPQHPPPAPSPAFDNLVYWDQ-----DP-----PERGAPPSTFKCTPT 1242
DB 1142 AQ-----PTCLSSGFSNPAWIKQSHQMSLNDPDYQDFFFKETKPNKIFKG-PT 1191
QY 1243 AENPEYLGLDVP 1254
DB 1192 AENAEYLRVAPP 1203

RESULT 4
QYH40
ID QYH40 PRELIMINARY; PRT; 1165 AA.
AC QYH40;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase proto-oncogene.
GN XMRK.
OS Xiphophorus xiphidium.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8086;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RX MEDLINE=98241172; PubMed=9582016;
RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RA Altschmid J., Schartl M.;
RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT overexpression and mutational alterations.";
RL Oncogene 16:1681-1690(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RA Schartl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53471; AADI0500.2; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC heme bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
```

| | | |
|----|-----------------------|--|
| SQ | SEQUENCE | 1165 AA; 129614 MW; 7F7BE38D8771A74E CRC64; |
| | Query Match | 40.4%; Score 2749; DB 13; Length 1165; |
| | Best Local Similarity | 45.8%; Pred. No. 1.7e-198; |
| | Matches | 584; Conservative 162; Mismatches 382; Indels 148; Gaps 29; |
| QY | 1 | MELAALCGWLLALLPPG-AAST---QVCTGTDKMLRLPASPETHDMLRHLVQGCVOV 55 |
| DB | 4 | LELEL-----LULLLSIGRCSTDPDRKVCQGTSNQMTM---LDNHYLKMKGKMGSCNV 56 |
| QY | 56 | VOGNLELYLPTNASLFLQDIQEVGYVLIHAHQVQVPLQRLRIVRGQTQLFEDNYALA 115 |
| DB | 57 | VLENLEITYTOENQDLSFLOSIOEVGYVLIAMNEVSTIPLVNLRLIRQNLIEGNTLL 116 |
| QY | 116 | VLDNGDPLNNQYIKANSK-----FIGITELQLSLTEILKGGVLIQRNPOLCYQDTILMKD 171 |
| DB | 117 | VMSN-----YQKNPSSPDVTVQVGLKQLSLNTEILSGGVKVSHPNLLCNVETINMWD 169 |
| QY | 172 | IFHKHKNQALALTLIDNRSACHPCSPMKGSCWGESSEDCQSLRTVVCAGGC-ARCKGP 230 |
| DB | 170 | IVDKTSNPTMNLIPHAERQCKQKOPGCVNGSCWAPGPHQKCFKLLCAEQNRCRGP 229 |
| QY | 231 | LPTDCHEOCAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290 |
| DB | 230 | KPIDCNEHCAGCTGPRATDCLACRDFNDGTCKDTCPPKXIYDIVSHQVVDNPNIKYT 289 |
| QY | 291 | FGASCVTACPNYLSTDVGSCTLVCLPHNQEVTAEDGTORCEKCKSPCARVCYGLGMEHL 350 |
| DB | 290 | FGAACVKECPSNVYVTE-GACVRSACAGMLEVD-ENGKRSCKPCDGVCPKVDGIGSL 347 |
| QY | 351 | REVRAVTSANIOEFAGCKKIFGSLAPLPESFGDGPASNTAPLOPQLOVFLEERTITGYL 410 |
| DB | 348 | SNTIAVNSNIGSFNCTKINGDIIILNRNSFGDPHYKIGPMDPEHMLNLTITVKEITGYL 407 |
| QY | 411 | YISAWPDSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWGLSLRSLRGSLALJH 469 |
| DB | 408 | VIMWMPENMTLSVQNLBEIRGRITFSRFGFVVVVQVSHLQWLGSLLSKEYSAGNVILK 467 |
| QY | 470 | HNTHLCFVHTVPWDQLFRNPQALLHTANRDECEVGEGLACHQICARGHCWGPPTQCV 529 |
| DB | 468 | NTPOLRYASTINWRRLFRSEDOSIEYDART-----ENQTCNNECEDGCGWPGPTMCV 520 |
| QY | 530 | NCSQFLRGQECVEECRLVQLPREYVNAHRHCLPCHPECOPONGSVTCFGEADQCVACAH 589 |
| DB | 521 | SLCHVDRGRCVASCNLQDPEPRAQVQDRCVQHQECLVQDTSITCYGPGPANGCKAH 580 |
| QY | 590 | YKDPFPCVACRSPGVKPDLSYMPIWKFDEEGACQPCINCTHSCVLDLDDKCPAEQAS 649 |
| DB | 581 | FQDGPQCIPRCPHGLMGDGTIL-INKYADKMGCQCPCQHCQCTQGCSPGLSGCRGD-IVS 638 |
| QY | 650 | PLTSIVSAVVGILLVVVLGVVFGIILKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQ 709 |
| DB | 639 | HSSLAVGLVSGLLITIVALLIWLRLRRRIK-RKRTIRLLQELVELVEPLTPSGQAPNQ 697 |
| QY | 710 | AQMRILKETELRKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILD 769 |
| DB | 698 | AFRLRLKETEFKDRVLGSGAGTVYKGLWNPDPENIRIPVAIKVLRNTSPKANKEILD 757 |
| QY | 770 | EAYVMAGVSPYVSRLLGICLTSTVOLVTQLMPYGCCLDVRHNRGRGLGSDQLLNWCMQI 829 |
| DB | 758 | EAYVMASVDHPHVCRLGICLTSAVOLVTQLMPYGCCLLDYVRHQHQRICGOWLLNWCQI 817 |
| QY | 830 | AKGMSYLEDLVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGCKVPIKWM 889 |
| DB | 818 | AKGMNYLEERHLVHRDLAARNVLLNPNHVKITDFGLSKLLTADEKEYQAHGCKVPIKWM 877 |
| QY | 890 | ALESILRRFRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTI 949 |
| DB | 878 | ALESILQWYTHQSDVWSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPQPICTI 937 |
| QY | 950 | DVTYIMVKWMIDSECRPRFELVSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLL 1009 |
| DB | 938 | EVTYIILKWMIDPSSRRPRFELVGEFSQMARDPYSRYLVIQ---NLSPSPDRRLFSRLL 994 |

| | | |
|--|---|--|
| QY | 1010 | EDDDMGDLVDABEYLVPOQGFCCPDAPAGAGGMVHRHSSSTRSGGDLTLGLEPSEE 1069 |
| DB | 995 | SSDD--DVVDADEYLL-----RYKRIN-RQS----- 1018 |
| QY | 1070 | APRSLAPSEGAGSDVFDGDLGMAKGLQSLPHTDPSPLORYSEDPV-PLPSETDGYV 1128 |
| DB | 1019 | -----EPCIPPNGH-----PVRENSIALRYISDPTQNALEKDLGDH- 1054 |
| QY | 1129 | APLTCSPQSPYVNVQDVRPOP-----PSPRE-----GPLP-AARPAGATLERAKTLSP 1175 |
| DB | 1055 | -----EYVNPQSETSSRLSDIYNPNYEDLTDGWPVSLSSQEAETNFSRPEYLNT 1105 |
| QY | 1176 | GKGVVGVDFAFEGGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLYWDDPPPERGAPPS 1235 |
| DB | 1106 | NQNSL---PLVSSGMDPDY---QAG-----YQAAF-----LPQTGALTG 1140 |
| QY | 1236 | TEKGTPTAENPEYLG 1251 |
| DB | 1141 | NGMFLPAAENLEYLG 1156 |
| RESULT 5 | | |
| Q9W6F6 | | |
| ID | Q9W6F6 | PRELIMINARY; PRT: 1137 AA. |
| AC | Q9W6F6; | |
| DT | 01-NOV-1999 | (TREMBLrel. 12, Created) |
| DT | 01-NOV-1999 | (TREMBLrel. 12, Last sequence update) |
| DT | 01-JUN-2002 | (TREMBLrel. 21, Last annotation update) |
| DE | Receptor tyrosine kinase (Fragment). | |
| GN | ERBB4. | |
| OS | Gallus gallus (Chicken). | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC | Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; | |
| OC | Gallus. | |
| OX | NCBI_TaxID=9031; | |
| RN | [1]_TaxID=9031; | |
| RP | SEQUENCE FROM N.A. | |
| RC | TISSUE=HINDRAIN; | |
| RX | MEDLINE=99263203; PubMed=10328884; | |
| RA | Dixon M., Lumsden A.; | |
| RT | "Distribution of neurogulin-1 (nrg1) and erbB4 transcripts in embryonic chick hindbrain." | |
| RL | Mol. Cell. Neurosci. 13:237-258(1999). | |
| DR | EMBL; AF121963; AAD31764.1; -- | |
| DR | HSSP; P11362; 1FGK. | |
| DR | InterPro; IPR000494; EGFR_L_domain. | |
| DR | InterPro; IPR000719; Euk_pkinase. | |
| DR | InterPro; IPR002174; Furin-like. | |
| DR | InterPro; IPR001368; TNFR_c6. | |
| DR | InterPro; IPR001245; Tyr_pkinase. | |
| DR | InterPro; IPR004019; YLP_motif. | |
| DR | Pfam; PF00757; Furin-like; 1. | |
| DR | Pfam; PF00669; pkinase; 1. | |
| DR | Pfam; PF01030; Recep_L_domain; 1. | |
| DR | Pfam; PF02757; YLP; 2. | |
| DR | PRINTS; PR00109; TYRKINASE. | |
| DR | ProDom; PD000001; Euk_pkinase; 1. | |
| DR | SMART; SM00261; FU_3. | |
| DR | SMART; SM00219; TyrKc; 1. | |
| DR | PROSITE; PS00107; PROTEIN_KINASE_ATP; 1. | |
| DR | PROSITE; PS00111; PROTEIN_KINASE_DOM; 1. | |
| DR | PROSITE; PS00109; PROTEIN_KINASE_TYR; 1. | |
| DR | PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1. | |
| KW | Kinase; Tyrosine-protein kinase. | |
| FT | NON_TER | 1 |
| SQ | SEQUENCE | 1137 AA; 12927 MW; 4D616436F87DC84F CRC64; |
| Query Match | | |
| Best Local Similarity | | |
| Matches | | |
| 539; Conservative 169; Mismatches 354; Indels 77; Gaps 24; | | |
| QY | 161 | LCYQDTILWKDIFHKHKNQALTLIDNRSRACHPCSPMKGSCWGESSEDCQSLRTVVC 220 |

| | | | |
|----|------|--|------|
| Db | 3 | LCFADTTHQDIVRNWASNFTLVPTNGSSGCGCHKSCGTG-RCWGPTENHCQTLTKTVC | 61 |
| Qy | 221 | AGGC-ARCKGPTDCCHCCAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTF | 279 |
| Db | 62 | AEQCDGRCYGVYVSDCHRECAGCGPKDTCFACNFNDSGACVTCQPTVYVPTTF | 121 |
| Qy | 280 | ESMNPPEGRYTFGASCVTACPNYLSVDGSCCTLVCPHLNQEVTAEDGTQRCCKSKPCA | 339 |
| Db | 122 | OLEHNNHAKYTYGAFCKVCKPHFV-VDSSSCVACRCPSSKMEV-EENGIMCKPCTDPC | 179 |
| Qy | 340 | RUCYGLGMEHLREVRVANTSANTQEPAGCKKIFGSLAFLPESFGDGPASNTAPLOQLQV | 399 |
| Db | 180 | KACDGTGTSLSAQTVDSSNDKFNCTKINGNLIFLVTGIRHGDYHTIAAINPEKLN | 239 |
| Qy | 400 | FETLEBITGYLYISAMPDLSVFQNLQVIGRILHMGAYSLTLQGLGISHLGRSLR | 459 |
| Db | 240 | FQTVREITGYLNTQSPENWTDPRVSNLVTIGRALYSGLSLILKQOIGITSLQFSLK | 299 |
| Qy | 460 | ELGSGALIHNTHLCTVHTVPMWDLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGH | 519 |
| Db | 300 | QISAGNIYITDNSLNCYVHTVNTSLFSTPSQKTVIHRNKAENCTADGMVCMNELCSSDG | 359 |
| Qy | 520 | CHGPGTQVCNCSQFLRGCEVCECVLQGLPREYVNAHCLPCHPECCP-QNGSVTCFG | 578 |
| Db | 360 | CMGPGDQCLSKCRFRIGRTCTIESCNLYDGEFREFANGSVCMCEDCQCEKMDNMITCVG | 419 |
| Qy | 579 | PEADQCVACAHYKDDPPFCVACRPSGVKPDLSYMPIWKFDEEGACOPCINCTHSCVDLD | 638 |
| Db | 420 | PGPDHCTKCFHFGDGNCEVCEKCPDGLGANSF--IFKYADEDECHPCPNCTQCGRGA | 477 |
| Qy | 639 | DKGC-----PAERASPLTSIVSAVV-GILLVVVLGVVFGILIKRROOKIRKYT | 686 |
| Db | 478 | SHDCIYVPTWQSTLPOHAR-TFL--IAAGVIGGLFIIVIMGLTFAVYVRKSIK-KKRA | 533 |
| Qy | 687 | MRLLOETELVEPLTSGAMPNOAQMRILKETELRKVKVLGSGAFGTGYKGIWPGENV | 746 |
| Db | 534 | LRRFL-ETELVEPLTSGTAPNAQRLILKETELRKVKVLGSGAFGTGYKGIWVPEGTV | 592 |
| Qy | 747 | KIPVAIKVIRENTSPKANKEILDEAYVMAGVGPYVSRLLIGLICLTSTVOLVTOLMPYGL | 806 |
| Db | 593 | KIPVAIKILNETTGPKANVEFMDALIMASMDPHLVRLLGVCLSTIQLVTLQMPHGCL | 652 |
| Qy | 807 | LHVRENRRGLSGDILLNMCQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGL | 866 |
| Db | 653 | LDYVHEHKDNGSOLLNMCVQIAKGMVLEERLVRDLAARNVLKSPNHVKITDFGL | 712 |
| Qy | 867 | ARLLDIDETEHADGKVPKMALESILRRRTHQSDVMSYGVYTWELMTFCAKPYDGI | 926 |
| Db | 713 | ARLEGDEKEYNADGKMPKMALECIHYRKFTHQSDVMSYGVYTWELMTFGKPYDGI | 772 |
| Qy | 927 | PAREIPDLLEKGERLPQPPICTIDVYIMVVKMIDSECRPRELVSEFSRWARDPQRF | 986 |
| Db | 773 | PTREIPDLLEKGERLPQPPICTIDVYIMVVKMIDADSRPKFKELAAEFSRWARDPQRY | 832 |
| Qy | 987 | VITQNEB-LGPASPLDSTFYRSLLEDMDGLVDABEYLVPOQGFPCPDAPAGAGGVHH | 1045 |
| Db | 833 | LVIQGDRLKPLSPNDSEKFNQLLDEEDLEMDABEYLV-PAENIPPIYSTRIDS | 891 |
| Qy | 1046 | RHRSSSTRSGGDLTLGLSESEAPRS--PLAP-SEGAGSVDFDGLGMGAAGLQSLP | 1102 |
| Db | 892 | NRNQFYVRDGGYAAEGV-PMPYRAPGCCIPEAPVAQGAETAFEDTCNGTLRKQVATL | 950 |
| Qy | 1103 | THDPSPLQRYSEDPYPLPS-----ETDGYVAPLTCSPQPEYVNPQDVRRPQPSREG | 1155 |
| Db | 951 | AKEDSSTORYSADPTVFIPERVIRGELDEDEGYTPMRDKPTDYLNPVENPVSRRKNG | 1010 |
| Qy | 1156 | PLPAA-RPAGATLERAKTISPGKQGVVQVDF-----AFGAVENPEYLTPOGGAAPQ | 1206 |
| Db | 1011 | DLQAVDNPEYH-----APNQPKAEDEYVNEPLYLNTFANTLENAEYL-----K | 1055 |
| Qy | 1207 | PHPPPAFSAFONLYYWDQDPPPERGA--PPSTFKGTPT-----AENPEYL | 1249 |

| | | | |
|----------|--|--|---------------|
| Db | 1056 | NNLPEKAKAFDNDPYMNHSLPPRSTLQHPDYLOEYSTKYFYKONGRIRDVAENPEYL | 1114 |
| RESULT 6 | | | |
| P79754 | | | |
| ID | P79754 | PRELIMINARY; | PRT; 1328 AA. |
| AC | P79754; | | |
| DT | 01-MAY-1997 | (TEMBLrel. 03, Created) | |
| DT | 01-MAY-1997 | (TEMBLrel. 03, Last sequence update) | |
| DT | 01-JUN-2002 | (TEMBLrel. 21, Last annotation update) | |
| DE | Bzbb3. | | |
| OS | Fugu rubripes (Japanese pufferfish) (Takifugu rubripes). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; | | |
| OC | Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; | | |
| OC | Tetraodontidae; Takifugu. | | |
| OX | NCBI_TaxID=31033; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=99177347; PubMed=10077531; | | |
| RA | Gellner K., Brenner S.; | | |
| RT | "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu | | |
| RT | rubripes."; | | |
| RL | Genome Res. 9:251-258 (1999). | | |
| DR | EMBL; AF056116; AAC34391.1; -. | | |
| DR | HSSP; P11362; 1FGK. | | |
| DR | InterPro; IPR000494; EGFR_L domain. | | |
| DR | InterPro; IPR000719; Euk_Pkinase. | | |
| DR | InterPro; IPR002174; Furin-like. | | |
| DR | InterPro; IPR001245; Tyr_Pkinase. | | |
| DR | Pfam; PF00757; Furin-like; 1. | | |
| DR | Pfam; PF00069; Pkinase; 1. | | |
| DR | Pfam; PF01030; Recep_L domain; 2. | | |
| DR | ProDom; PD000001; Euk_Pkinase; 1. | | |
| DR | SMART; SM00261; FU; 3. | | |
| DR | SMART; SM00219; TyrKC; 1. | | |
| DR | PROSITE; PS00107; PROTEIN KINASE ATP; 1. | | |
| DR | PROSITE; PS00011; PROTEIN_KINASE_DOM; 1. | | |
| KW | ATP-binding; Transferase. | | |
| SK | SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64; | | |
| Qy | 9 | WGLLALLPP--GAASTQ---VCTCTDMKRLRASPETHLDMRLHLYQCGOVQGNLEL | 62 |
| Db | 4 | WRLILMCVASRLRAASSQTQAEVCPGTQNGLSSTGSOENQYNLKDORYKGEIIMGNLEI | 63 |
| Qy | 63 | TYLPTTNASLSFLQDIOEVQYVLIHQNVRQVPLQRLRIVRGTLQFEDNYALAVLDNGDP | 122 |
| Db | 64 | TOIESNWFSLKTIREVTVGVLIANNHFOEIPLQGLRVIRGNSLYERRFALSFLN--- | 120 |
| Qy | 123 | LNNQYTKANSKFTIGTELQRLSLEILKGGVLIFQNPQLCYQDTILWKDIFHKNNQAL | 182 |
| Db | 121 | ---YKPDGFS--GLNQLGLMNLTEILDGVQIINNLYRYPGWVYWRDII-RNNDAPIE | 173 |
| Qy | 183 | LIDTNRSRACHPCSPMKCSRCSGESSEDCQSILTRTVACGCC-ARCKGPLPTCCHEQCA | 241 |
| Db | 174 | IOFNGERGVCYH--KSC-GNYCMGPGKQDQQLTKTVCAQCNCNDRFCGTSPRCCCHIEA | 229 |
| Qy | 242 | AGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPY | 301 |
| Db | 230 | ACGKGPLDTCFACRLFNDSGACVPCPQTLIYNKQTFQWETNPNAKYQVGSICVQCPT | 289 |
| Qy | 302 | NYLSTDVSGCTLVCPHLNQEVTAEDGTQ-CEKSKPCARVCYVGLGMEHLREVRVANTS | 360 |
| Db | 290 | HFV-VDSGSCSVCPDKMEV--ERGSQRCCLCSGLCPKVCBGTGAE---QRQTVSSN | 343 |
| Qy | 361 | IQEFACKKIFGSLAFLPESFGDGPASNTAPLOPELOVPELEBITGYLYISAWDSLP | 420 |
| Db | 344 | IDSFINCTIQGSLHFLVTGILGDDPKNPVPPDLDAKKLEVFRVREITDILNQSMPKELN | 403 |

Query Match 34.0%; Score 2315; DB 13; Length 1328;
Best Local Similarity 40.6%; Pred. No. 1.2e-165;
Matches 522; Conservative 155; Mismatches 419; Indels 190; Gaps 31;

```
QY 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGSIWGLRLSRLSGSLALIHHTHLCFVHTV 480
Db 404 DLSVFSSLTQGRSLFKRFLSMVMRIPTLTSLGLSLREISDGSYVISQNAHLCHVHTV 463
QY 481 PWDQLFRNPH-QALLHTANRPEDECVEGLACHOLCARGHCWGPQTQCVNCSQFLRGQE 539
Db 464 NWTQLFRGRVRANSLSNRPAECVADRVCPLCSDSGCGWGPDPQCLSCRNYSRHT 523
QY 540 CUEECRVLOGLPREVYNARH-CLPCHPEQOPQNGSVTCFPGPADQCVACAHYKDPFPCA 598
Db 524 CVAGCHFNSGIPREFAGLNGVCVACHPECKPQTGKASCTGPGADECMACTKFRDGPYCMS 583
QY 599 RCPGSGVKPDLSPYIWKFPDEGACOPCPINCTHSCVDLDDGCPAERASPLTSUSAV 658
Db 584 SCPAGVN-DGEKGLIFKFPNRECHCEPCQNCTQCGSGPLNDC-----LEAARLTSSGOI 639
QY 659 VGILLVVLGVVVF-----GILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNOAQ 711
Db 640 TGIALGVPAGLIFCLVFLGLMYHRLAIRKRAMRYLSEGESFEPLGP-GEKGTKVH 698
QY 712 MRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 771
Db 699 ARILKPSDLRKIKPLGSGVFGTVSGFMIPGETVKIPVAIKTIQDSSGRQTFTEITDHL 758
QY 772 YVMAGVGSYPVSRLLGICLTSTVQLVTQLMPCCLLDHYENRGRGLSGODLLNWCQIAK 831
Db 759 LSMGSLDHPYIVRLGICPGTCLQVTLQSSHGSLEHROHKTSLDORLLNWCVOIAK 818
QY 832 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEYHADGCKVPKIKMAL 891
Db 819 GWYILEHRVHVHKNLAARNILKNDYQVQISDYGVADLLYPDDKKVYSETKTPKIKMAL 878
QY 892 ESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDV 951
Db 879 ESILFRYTHQSDVMSYGVTVWEMMSFGAEPYASVQPEVPSVLEKGERLSQPAICTIDV 938
QY 952 YMIWVKMIDSECPREFELVSESRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLED 1011
Db 939 YMWVKMIDENIRPTFKELASDFTRMARDPRYLVRMEG-----ED 982
QY 1012 DMGDVDAEYLVPOQGFCCPDAPAGAGMVHRRHSSTRSGGDLTLGLEPSEEEAP 1071
Db 983 SCMGFEFL-----RGCSE-GLLEADLEEDEE- 1008
QY 1072 RSLAPSGAGSDVFDGLMG----AAKGLSLTHDPSPLQ-----RYSEDPTV 1118
Db 1009 -----GLDRFATPSLQSPSPSWSTSPSINSVMVMTQLRYD----- 1044
QY 1119 PLPSETDGVAPLTCSP-EPVNO-----PDVRRPQPPSPREGPL--PAAR 1161
Db 1045 -FAVSQGGHIGVLPSPSPVDTIQMLWYORSLSVRTLPRSAFRRSREAELEDGAG 1103
QY 1162 PAGATLERAKTLSPGKNGVYKDVAFAGGAVERNPEYLTPOGGAAPQHPPPAPSPAFDNL 1221
Db 1104 CAGIFRVR-----FGSERGN-----POGG----- 1122
QY 1222 YWDQPPPERGAPPSTFKGTPTAENPE 1247
Db 1123 --QQRKLSTASSPSSFKTWADEDE 1146
```

RESULT 7

```
Q9BIH9 PRELIMINARY; PRT: 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
```

```
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SUJ;
RA Lyett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
epidermal growth factor receptor."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1;
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER 1
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D8967724F07 CRC64;
```

```
Query Match 30.1%; Score 2047.5; DB 5; Length 1433;
Best Local Similarity 32.7%; Pred.No. 2.2e-145;
Matches 473; Conservative 196; Mismatches 393; Indels 385; Gaps 36;
```

```
QY 26 CTGTDMLKRLPASPETHLDMLRHLQYGCQVQVQCNLELYLPTNASLSFLQDIOEVQGYVL 85
Db 1 CIGTNGRMSVPANREHYKYLDRYTNCTYVDGNLEITWIONITDLNFLOHIREVTGYVL 60
QY 86 IAHNVQVPLQRLRIVRGTQLF-----EDNYALAVLDNGDPLNNYIKANSKFIQITEL 140
Db 61 ISLYLDPQVILPRLQIIRGRTTFKLKNKEEAYGLFV-----SFSHWNTL 104
QY 141 QLRSLTEILKGVLLQRLNOLCVQDTILWKDI-FHKNNQLALTLDITNRSRACHPCSPMC 199
Db 105 ELPALRDLILGSGVGFNFNNLCHMKSIINWEEILLAPQTSMQYTFNFSSPERVCPCHPSC 164
QY 200 KGSRCWGESSEDCQSLTRTVACAGCA--RCKGPLPTDCCHCCCAAGCTGPKHSDCLACLH 257
Db 165 EVG-CWGEAHNCQRFSLNCSPOCSQGRCFGPKPRECHLFCAGGCTGTQSDCLACKN 223
QY 258 FHSIGICELHCALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTVDVSGCTLVCPPL 317
Db 224 FYDDGVCKQCECPPMQIYNPTNYFWEPNPDGKYATGATCVRKCP-EHLKXDNACAVRKCPK 282
QY 318 HMQEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRAVTSANIQEFACKKIFGSLAFL 377
Db 283 GKMPQNSE-----CVPCKGVCPKTCPEGIVH-----SNIGNYKDKCTTIEGSLLEIL 329
QY 378 PESFDGDPASNT-----APLPEQLQVFTLEBEITGYLYISAWPDSLPLDSVFQNLQ 429
Db 330 DQSFDFGQQVYTNFSFGPRYIKIDPDRLEVFSTVKEITGFINIAHHHPNFTLTLYNFRNLE 389
QY 430 VIRGILHNGAY-SLTLOGLGISWGLRLSRLSGSLALIHHTHLCFVHTVVPDOLFNR 488
Db 390 VVGGRLKENLFASVYIVKTSLSKLSLEKSLKRVNSGSIIVLENSDLCFVEDIDWSIKKS 449
QY 489 PHQALLHTANRPEDECVEGLACHOLCARGHCWGPQTQCVNCSQFLRQECVEECRVLQ 548
Db 450 SDHEVWVQKNRNATECHEEGMECECSQCSKAGCKGKGPQCEUCKNVKYKGKCLDSCK--- 506
```



```
DR EMBL; BC027080; AAH27080.1; -.
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match      25.5%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 5.7e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICT 948
Db 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICT 60

QY 949 IDVYIMVWKWIDSECRPRFRELSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1008
Db 61 IDVYIMVWKWIDSECRPRFRELSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120

QY 1009 LEDDDMGDLVDAEYLVPOOGFCPPAPGACGQVHRRSSSTRSGGGLTLGLEPSEE 1068
Db 121 LEDDDMGDLVDAEYLVPOOGFCPPALGTGSTAHHRRSSSARSGGGLTLGLEPSEE 180

QY 1069 EAPRSPAPSEAGSDVDFDGLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYV 1128
Db 181 EAPRSPAPSEAGSDVDFDGLAVGVTKGLQSLSPHDLSPLOQYSEDPTVPLPSETDGYV 240

QY 1129 APLTCSPOQYVNPQDVRPQSPREGPLPAARPAGATLERAKTUSPGKNGVVKVDFAFG 1188
Db 241 APLTCSPOQYVNPQDVRPQSPREGPLPAARPAGATLERAKTUSPGKNGVVKVDFAFG 300

QY 1189 GAVENPEYLTPOGGAPOPHPAFSPAFDNLYYNDODPPEGAPSTFTGPTAENPEY 1248
Db 301 GAVENPEYLTPOGGAPOPHPAFSPAFDNLYYNDODPPEGAPSTFTGPTAENPEY 360

QY 1249 LGLDVPV 1255
Db 361 LGLDVPV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF02813; Retro_M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;

Query Match      25.3%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 4.3e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 569 PONGSVTCFPGPADQCVACAHYKDPFCFVARCPSGVKPDLSYMPIWKFPDEEGACQCPPI 628
Db 141 PBEATPTKTPG--DHCMKCAHFDGPHCVKACPAVGLGENDTL-VMKYADANAVCQLCHP 197

QY 629 NCTHSCVDLDDKGCAPAEQASPLTSTVSAAV--GILLAVVLGVVFGILLIKRQOKIRKYTE 687
Db 198 NCTRCKGPGLEGGCP---NGSKTPSIAAGVVGGLCLVVGUGIGLYLRRR--HIVKRTL 253

QY 688 RRLLOETELVELTPSGAMPNOAQRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVK 747
Db 254 RRLQERELVELTPSGEAPNOAHLRLAKETEFKKVKVLGSGAFGTVYKGLWIPGEKVK 313

QY 748 IPVAIKVLRENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLL 807
Db 314 IPVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQMLPYGCLL 373

QY 808 DHVRENRRGLSGSODLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDPLA 867
Db 374 DVIHKONIGSQYLLNWCQIAKGNVLEERLVRDLAARNVLKTPHVKITDPLA 433

QY 868 RLLDIDETEHADGGKVPKMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI 927
Db 434 KLGADKEHYHAEGKVPKMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI 493

QY 928 AREIPDLLEKGERLPQPPICTIDVYIMVWKWIDSECRPRFRELSEFMRMARDPQRFV 987
Db 494 ASEISVLEKGERLPQPPICTIDVYIMVWKWIDADSRPKRELIAEFSKWARDPPRYL 553

QY 988 VIQ-NEDLGPASPLDSTFYRSLLEDGDLVDAEYLVPOOGFCPPAPGAGGVVHHR 1046
Db 554 VIQGDREHPLPSTDSKFYRTLMEBEDMEDIVDADSYLVPHOGFF----- 598

QY 1047 HRSSSTRSGGDLTLGLEPSEEEAPRSP---APSEAGSDVDFDGLGMAAGLQSL 1101
Db 599 -NSPST-----SRTPLLSLSATSNNSATNCID-----RNGQGH 631

QY 1102 PTHDPSPLQRYSEDPTVPLPSET--DGYVAPITCSGPQYVNPQDVRPQSPREGPLPA 1159
Db 632 PVREDSFVQRYSSDPTGNFLESIIDGFL-----PAPEYVQ--LMPKKPS----- 675

QY 1160 ARPAGATLERAKTUSPGKNGVVKDFP-----AFGGAVENPEYL 1197
Db 676 -----TAMVQNIYNNISLTAISKLPMDSRVYQNSHSTAVDNPEYL 715

RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE V-erbB protein (fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
```

```
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; kinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00219; TykKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR Tyrosine-protein kinase.
DR NON TER
FT
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF25E1 CRC64;

Query Match 25.2%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 4.2e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

Qy 578 GPEADQCVACAHYKDPFCFCVAPCPGKVPDLSYMPKFPDEEGACQPCPINCTHSCVDL 637
Db 1 GP--DHCKCAHFIDGPHCVKACGAGVLGENDTL-VMKYADANAVCOLCHNCPTCKGCP 57
Qy 638 DDKGCPAEQASPLTSIVSAV-GILLVVLGVVFGILIKRQOKIRKYTMRLLOQETEL 696
Db 58 GLEGCP---NGSKTPSTAAGVVGGLCLLVVGLGIGLYLRR-HIVKRTLRLRLQREL 113
Qy 697 VEPLTPSGAMPNOAQMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNOAHLKETETFKVKVLGSGAFGVYKGLWIPGEKVKIPVAIKELR 173
Qy 757 ENTSPKANKEILDEAYVMAGVGSYPVSRLLGICLTSTVOLVTQMLPYGCLLDHVRENRR 816
Db 174 EATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQITQMLPYGCLLDYIREHKN 233
Qy 817 LGSQDLNWCQIAKMSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLIDETE 876
Db 234 IGSQDLNWCQIAKMSYLEDVRLVHRDLAARNLVKTPQHVKITDFGLAKLLGADEKE 293
Qy 877 YHADGGKVPKWALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 294 YHAEGGKVPKWALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISVLE 353
Qy 937 KGERLPPOPICTIDVYIMVWKWMDSECRPRFRELVSFESRMARDPQRFVWIQ-NEDLG 995
Db 354 KGERLPPOPICTIDVYIMVWKWMDADSRPKRELIAEFSKWARDPPRYLVVLCQDERMH 413
Qy 996 PASPLDSTFYRSLLDDMDGLVDAEYLVPOQGFCCPDPAAGAGMWHHRSSSTRSG 1055
Db 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTLCLEPSEEEAPRSL-----APSEGAGSDVFDGLMGAAKGLQSLTHDPSPLQ 1110
Db 455 -----SRTPLLSLSATSNNSATNCID-----RNGOGHPVREDSEFVQ 491
Qy 1111 RYSEDPVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARAGATLE 1168
Db 492 RYSDPTGNFLRESIDGFL-----PAPEYVNO--LMPKKPS----- 526
Qy 1169 RAKTSPGKGVKVDVF-----AFGGAVENPEYL 1197
Db 527 ----TAMVQIQYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 566

RESULT 12
Q8WTV0
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN PP3659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF318349; AAL55856.1; -.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; kinase; 1.
DR Pfam: PF02757; YLP; 2.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00219; TykKc; 1.
DR PROSITE: PS00018; EF HAND; UNKNOWN 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FE3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 9.2e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

Qy 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 948
Db 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60
Qy 949 IDVYIMVWKWMDISECRPRFRELVSFESRMARDPQRFVVIQNEIDLGPAFLDSTFYRSL 1008
Db 61 IDVYIMVWKWMDISECRPRFRELVSFESRMARDPQRFVVIQNEIDLGPAFLDSTFYRSL 120
Qy 1009 LEDDDMGDLVDAEYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGLDLCLEPSEE 1068
Db 121 LEDDDMGDLVDAEYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGLDLCLEPSEE 180
Qy 1069 EAPRSPALASEGAGSDVFDGLMGAAKGLQSLTHDPSPLQRYSEDPVPLPSETDGYV 1128
Db 181 EAPRSPALASEGAGSDVFDGLMGAAKGLQSLTHDPSPLQRYSEDPVPLPSETDGYV 240
Qy 1129 APLTCSQPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTSLGKGVKVDVFAFG 1188
Db 241 APLTCSQPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTSLGKGVKVDVFAFG 300
Qy 1189 GAVENPEYLTPOGGAAPQ-----HPPPA---FSPAFDNL 1220
Db 301 GAVENPEYLTPOGGAALSPTLLPSAQPSTTITGTRTHQSGGLHPAPSKGHLRQRTQST 360
Qy 1221 YVWD-QDPPER-----GAPPSTFKGTPTAEN 1245
Db 361 WWTCCOPEGQVRRSPDVSSSGSREGLTSAKIKRWEGPPTISRGTCHARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
DE Gag, v-erb-A, v-erb-B.
GN PP3659.
OS Avian erythroblastosis virus.
OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90206603; PubMed=1969616;
RA Bruskina A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";
```

RL Oncogene 5:15-24(1990).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk pkinaase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Strhmn_receptor.
DR InterPro; IPR001245; Tnf_pkinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; Zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; TyrKc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 24.3%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 6.9e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 541 VEECRVLQGLPRE-YVNAR-HCLP-----CHPEQC 568
Db 354 IEKQESYLLAFEHYINVRKNIHPFWSKLLMKVADLRMIGAYHRSFLHMKVCEPTELS 413
QY 569 PONGSVTCFGEADQVACAHYKDPFCVACPSGVKPDLSYMTWKPDDEGACQPCPI 628
Db 414 PQE-----VGP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHP 465
QY 629 NCTHSCVDLDDKCPAEQORASPLTSIVSAVV-GILLVVVLGVVFGVILIKRQOKIRKYTM 687
Db 466 NCTRCCKGPGLEGCP--NGSKTPSIAGVVGGLLCLVVGIGLYLRRR-HIVRKTL 521
QY 688 RRLQOETELVPLTPSGAMPNOAQRILKETELRKVKVGLGSAFGTVYKGIWIPDGENVK 747
Db 522 RRLQOERELVPLTPSGEAPNOAHLRLKETEFKKVKVGLGFGAGTVYKGLWIPEGEKVT 581
QY 748 IPVALKVLRENTSPKANKEILDEAYVMAGVGSPPVSRLLGICLTSTVOLVTQLMYPGCLL 807
Db 582 IPVALKEUREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLITQLMYPGCLL 641
QY 808 DHVRENRRGLASQDLLNMCQIAGMSYLEDVRLVHRDLAARNVLVKSPPNHVKITDFGLA 867
Db 642 DYIREHKDNGISQYLLANCVCQIAKGMNLYEERHVRDLAARNVLVKTPOHVKITDFGLA 701
QY 868 RLLDIDETEHADGKVPDKMALESILRRRPTHQSDVMSYGVTVVWELMTFGAKPYDGIP 927
Db 702 KQLGADEKEYHAEGKVPDKMALESILHRIYTHQSDVMSYGVTVVWELMTFGSKPYDGIP 761
QY 928 AREIPDLLEKGERLPQPPICITDVYIMVWKMDSECRPFRELVSFESRMARDPQRFV 987
Db 762 ASEISSVLEKGERLPQPPICITDVYIMVWKMSGADSRPFRELIAEFSEKMDRPPRYL 821
QY 988 VIQ-NEDLGPASPLDSTFYRSLLEDVDAEYLVPOQGFPCDPAPGAGMGVHR 1046
Db 822 VIQGERMHLSPSTDSKYRFLMEEDMEDIVDAEYLVPHQGF----- 866
QY 1047 HRSSTSTRGGGDLTLGLPEPSEEAAPRSLAPSEAGSDVFDGLMGAAKGLQSLPTHDP 1106
Db 867 -NSPST-----SRTPLLSSLATS-----NSATKCIDRNGGH-- 898

QY 1107 SPLQRYSEDPVTPLPSETDGYVAPLTCSQPQBYVQNPQVVRPQPSREGPLPAARPAGAT 1166
Db 899 -----EVREDGFL-----PAPEYVQ--LMPKKPSTAMVQNOIYNYISLT 936
QY 1167 -LERAKTILSPGKNGVVKDVFAFGGAVENPEYL 1197
Db 937 AISKLPMSRYQN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CAA30024.1; -.
DR HSSP; P11362; 1FCG.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CCA0F8AF4 CRC64;

Query Match 24.2%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 1.3e-115;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPEADQCVACAHYKDPFCVACPSGVKPDLSYMPIMKFPDDEGACQPCPINCTHSCVDL 637
Db 1 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHPNTRCCKGP 57
QY 638 DDKCPAEQORASPLTSIVSAVV-GILLVVVLGVVFGVILIKRQOKIRKYTMRLLOETEL 696
Db 58 GLEGCP--NGSKTPSIAGVVGGLLCLVVGIGLYLRRR-HIVRKTLRLLQEREL 113
QY 697 VEPLTPSGAMPNOAQRILKETELRKVKVGLGSAFGTVYKGIWIPDGENVKIPVALKVL 756
Db 114 VEPLTPSGEAPNOAHLRLKETEFKKVKVGLGFGAGTVYKGLWIPEGEKVTIPVALKEUR 173
QY 757 ENTSPKANKEILDEAYVMAGVGSPPVSRLLGICLTSTVOLVTQLMYPGCLLDHVRENRR 816
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLITQLMYPGCLLDYIREHKD 233
QY 817 LGSQDLLNMCQIAGMSYLEDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNMCVQIAKGMNLYEERHVRDLAARNVLVKTPODKITDFGLAKQLGADEKE 293
QY 877 YHAGGKVPDKMALESILRRRPTHQSDVMSYGVTVVWELMTFGAKPYDGIPAREIPDLLE 936
Db 294 YHAGGKVPDKMALESILHRIYTHQSDVMSYGVTVVWELMTFGSKPYDGIPASEISSVLE 353
QY 937 KGERLPQPPICITDVYIMVWKMDSECRPFRELVSFESRMARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPPICITDVYIMVWKMSDADSRPFRELIAEFSEKMDRPPRYLVIOGDERMH 413

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-103-117-12

Perfect score: 6809

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq.101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|------------------------------|
| 1 | 6736 | 98.9 | 1255 | 21 | AA1992620 Human heregulin 2 |
| 2 | 6736 | 98.9 | 1255 | 22 | AAE12130 Human tyrosine kin |
| 3 | 6736 | 98.9 | 1255 | 22 | AAB60167 HER2 transgene pla |
| 4 | 6736 | 98.9 | 1255 | 23 | AAU74545 Human HER2 (ErbB2) |
| 5 | 6730 | 98.8 | 1255 | 17 | AAW01111 HER-2/neu protein. |
| 6 | 6730 | 98.8 | 1255 | 20 | AAW92406 Human HER-2/neu on |
| 7 | 6730 | 98.8 | 1255 | 21 | AAB21198 Human HER-2/neu pr |
| 8 | 6730 | 98.8 | 1255 | 21 | AA1984780 Amino acid sequenc |
| 9 | 6730 | 98.8 | 1255 | 22 | AAB85458 Human HER-2/neu pr |
| 10 | 6730 | 98.8 | 1255 | 22 | AAG88267 HER2/neu amino aci |

| | | | | | |
|----|--------|------|------|----|------------------------------|
| 11 | 6730 | 98.8 | 1255 | 23 | AAE24067 Human Her-2 protei |
| 12 | 6730 | 98.8 | 1255 | 23 | AAE20479 Human Her-2/neu pr |
| 13 | 6730 | 98.8 | 1255 | 23 | AAAM51143 Human Her-2/neu on |
| 14 | 6730 | 98.8 | 1255 | 23 | AAU77114 Human Her-2/neu po |
| 15 | 6687 | 98.2 | 1433 | 14 | AAR39568 Sequence of c-erbB |
| 16 | 6566 | 96.4 | 1223 | 23 | AAU98923 Human breast cance |
| 17 | 6413 | 94.2 | 1200 | 21 | AAB21208 Human HER-2/neu pr |
| 18 | 5967.5 | 87.6 | 1256 | 21 | AAB21199 Rat Her-2/neu prot |
| 19 | 5967.5 | 87.6 | 1256 | 23 | AAAM51144 Mouse Her-2/neu on |
| 20 | 5939.5 | 87.2 | 1256 | 21 | AAB21206 Mouse Her-2/neu pr |
| 21 | 5939.5 | 87.2 | 1256 | 22 | AAG62860 Amino acid sequenc |
| 22 | 5939.5 | 87.2 | 1256 | 23 | AAAM51151 Mouse Her-2/neu on |
| 23 | 4816 | 70.7 | 919 | 21 | AAB21203 Human HER-2/neu fu |
| 24 | 4089.5 | 60.1 | 926 | 23 | AAAM51148 Her-2/neu extracel |
| 25 | 4089.5 | 60.1 | 926 | 23 | AAAM51152 Mouse Her-2/neu ex |
| 26 | 4089.5 | 60.1 | 926 | 23 | AAAM51153 Mouse Her-2/neu ex |
| 27 | 3700 | 54.3 | 712 | 21 | AAB21204 Human HER-2/neu fu |
| 28 | 3700 | 54.3 | 712 | 23 | AAAM51149 Her-2/neu extracel |
| 29 | 3554 | 52.2 | 782 | 18 | AAW19764 Her2-QM-CSF immuno |
| 30 | 3552 | 52.2 | 653 | 21 | AAB21200 Extracellular HER- |
| 31 | 3552 | 52.2 | 653 | 23 | AAAM51145 Human Her-2/neu on |
| 32 | 3514 | 51.6 | 645 | 22 | AAB60408 Human ErbB2 oncopr |
| 33 | 3514 | 51.6 | 645 | 22 | AAB61593 Human ErbB2 extrac |
| 34 | 3449 | 50.7 | 951 | 21 | AA144933 DC9CFV-erbB2EC fu |
| 35 | 3349 | 49.2 | 624 | 11 | AAAR08222 Extracellular port |
| 36 | 3160 | 46.4 | 1210 | 21 | AAB19259 Amino acid sequenc |
| 37 | 3160 | 46.4 | 1210 | 21 | AA19259 Human EGF receptor |
| 38 | 3160 | 46.4 | 1210 | 23 | AAE23019 Human Her-1 protei |
| 39 | 3160 | 46.4 | 1210 | 23 | AAAM50768 Human epidermal gr |
| 40 | 3158 | 46.4 | 1210 | 22 | AAB68420 Amino acid sequenc |
| 41 | 3119 | 45.8 | 1210 | 23 | ABP51768 Human epidermal gr |
| 42 | 3084 | 45.3 | 583 | 23 | AAE20481 Human protein for |
| 43 | 3084 | 45.3 | 587 | 23 | AAE20481 Human protein for |
| 44 | 3083 | 45.3 | 589 | 23 | AAE20484 Human protein for |
| 45 | 3083 | 45.3 | 600 | 23 | AAE20482 Human protein for |

ALIGNMENTS

RESULT 1
AA1992620
ID AA1992620 standard; Protein; 1255 AA.
XX AC AA1992620;
XX AC
XX 10-AUG-2000 (first entry)
XX DE Human heregulin 2 (Her2).
XX KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX OS Homo sapiens.
XX FH Key
FH Domain
FT FT Location/Qualifiers
FT FT 1..173
FT FT /label= N-terminal
FT FT /note= "mature polypeptide"
FT FT 5..25
FT FT /label= insertion region
FT FT /note= "suitable for foreign epitope insertion"
FT FT 59..73
FT FT /label= insertion region
FT FT /note= "suitable for foreign epitope insertion"
FT FT 103..117
FT FT /label= insertion region
FT FT /note= "suitable for foreign epitope insertion"
FT FT 149..163
FT FT /label= insertion region
FT FT /note= "suitable for foreign epitope insertion"
FT FT 174..323

FT Region /label= Cysteine_rich_domain
FT 210..224
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT 250..264
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Domain
FT 324..483
FT /label= Ligand_binding_domain
FT Region
FT 325..339
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT 369..383
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT 465..479
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Domain
FT 484..623
FT /label= Cysteine_rich_domain
FT Region
FT 579..593
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Domain
FT 624..654
FT /label= Transmembrane_domain
FT Region
FT 632..652
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT 653..667
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Domain
FT 655..1010
FT /label= Tyrosine_kinase_domain
FT Region
FT 661..675
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT 695..709
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT 710..730
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Domain
FT 1011..1235
FT /label= C-terminal_domain
FT XX
PN WO200020027-A2.
XX
XX 13-APR-2000.
XX
XX 05-OCT-1999; 99WO-DK00525.
XX
XX 05-OCT-1998; 98DK-0001261.
XX 20-OCT-1998; 98US-0105011.
XX
XX (MEBI-) M & E BIOTECH AS.
XX
XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
XX Gautam A, Birk P, Karlsson G;
XX WPI: 2000-349917/30.
XX N-PSDB; AAA09455.
XX
XX Inducing immune responses to weakly immunogenic, tumor associated
XX peptide antigens for the treatment of breast and prostate cancer
XX
XX Claim 62; Page 193-198; 220pp; English.
XX
XX This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
XX Her2 can be used in the claimed method as an autovaccine to induce a CTL
XX response. Subdominant CTL epitopes, antibody binding regions and
XX cysteine residues involved in disulfide bonds are preserved in the
XX immunogenized forms. Regions suitable for the insertion of foreign T
XX helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
CC cell-associated peptide antigens (PA) such as those associated with
CC cancers (self-proteins), e.g. human prostate specific membrane antigen
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC The method comprises effecting simultaneous presentation by antigen
CC producing cells (APCs) of the animals immune system of: (1) at least 1
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC B-cell group derived from the cell-associated PA; and (2) at least 1
CC first T-helper cell group which is foreign to the animal. Analogues of
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC part of all known and predicted CTL and B-cell epitopes of the respective
CC PA and including at least one foreign T helper epitope are also claimed.
CC The method is used to treat prostate, prostate/breast or breast cancer
CC when the PA is human PSM, FGF8b and Her2, respectively.
XX
XX Sequence 1255 AA;

| | Query Match | Best Local Similarity | 98.9%; | Score 6736; | DB 21; | Length 1255; |
|----|---------------|---|--------|-------------|--------|--------------|
| | Matches 1243; | Conservative | 2; | Mismatches | 10; | Indels |
| | | | | | 0; | Gaps |
| QY | 1 | MELAAALCRWGLLLALLPPGAASTQVCTGTDKMLRLPASPETHLDMRLHLYQGVVQGNL | 60 | | | |
| DB | 1 | MELAAALCRWGLLLALLPPGAASTQVCTGTDKMLRLPASPETHLDMRLHLYQGVVQGNL | 60 | | | |
| QY | 61 | ELTYLPTNASLSFLQDIOEVQGYVLIHNNQVQVPLQRLIRIVRGTLQFEDNVALAVLDNG | 120 | | | |
| DB | 61 | ELTYLPTNASLSFLQDIOEVQGYVLIHNNQVQVPLQRLIRIVRGTLQFEDNVALAVLDNG | 120 | | | |
| QY | 121 | DPLNNQYIKANSKFITELQRLSLTEILKGGVLIQPNPOLCYQDITLWKDIFHKNNQLA | 180 | | | |
| DB | 121 | DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQPNPOLCYQDITLWKDIFHKNNQLA | 180 | | | |
| QY | 181 | LTLDITNRSRACHPCSPMKSGRCWSESSEDQSLTRTVCAAGCARCKGPLETDCHEQC | 240 | | | |
| DB | 181 | LTLDITNRSRACHPCSPMKSGRCWSESSEDQSLTRTVCAAGCARCKGPLETDCHEQC | 240 | | | |
| QY | 241 | AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESNPNEGRTVTFGASCVTACP | 300 | | | |
| DB | 241 | AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESNPNEGRTVTFGASCVTACP | 300 | | | |
| QY | 301 | YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN | 360 | | | |
| DB | 301 | YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN | 360 | | | |
| QY | 361 | IOEFAGCKKIFGSLAFLPESPDGDPASNTAPLOPEQLQVFETLEITGYLISAMPDLSL | 420 | | | |
| DB | 361 | IOEFAGCKKIFGSLAFLPESPDGDPASNTAPLOPEQLQVFETLEITGYLISAMPDLSL | 420 | | | |
| QY | 421 | DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELCSGLALIHNNTHLCFVHTV | 480 | | | |
| DB | 421 | DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELCSGLALIHNNTHLCFVHTV | 480 | | | |
| QY | 481 | PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSOFLRGQEC | 540 | | | |
| DB | 481 | PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSOFLRGQEC | 540 | | | |
| QY | 541 | VEECRVLOGLPREYVNAHCLPCHPECCQPNQSVTCFGEADQCVACAHYKDPFCVARC | 600 | | | |
| DB | 541 | VEECRVLOGLPREYVNAHCLPCHPECCQPNQSVTCFGEADQCVACAHYKDPFCVARC | 600 | | | |
| QY | 601 | PSGVKPDLSYMPIWKFPDDEGACQPCINCTHSCVDLDDKGPAPQORASPLTSIVSAVVG | 660 | | | |
| DB | 601 | PSGVKPDLSYMPIWKFPDDEGACQPCINCTHSCVDLDDKGPAPQORASPLTSIVSAVVG | 660 | | | |
| QY | 661 | ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL | 720 | | | |
| DB | 661 | ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL | 720 | | | |
| QY | 721 | RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEATVMAGVGP | 780 | | | |
| DB | 721 | RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEATVMAGVGP | 780 | | | |

QY 781 YVSRLLGICLTSTVOLTPQMPYGCCLLDHVENRGRGLGSQDILLNWCMIAGKHSYLEDVR 840
DB 781 YVSRLLGICLTSTVOLTPQMPYGCCLLDHVENRGRGLGSQDILLNWCMIAGKHSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKPKWMALESILRRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKPKWMALESILRRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
QY 961 IDECEPRPRELVSEFSRMARDPQRFVIONEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
DB 961 IDECEPRPRELVSEFSRMARDPQRFVIONEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
QY 1021 EYLVLPQOGFFCDDPAPGAGMVHRRHSSTSGGGDLTLGLEPSEEEAPRSLAPSEG 1080
DB 1021 EYLVLPQOGFFCDDPAPGAGMVHRRHSSTSGGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVREPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1200
DB 1141 NQPDVREPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
DB 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2

AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
DT 18-DEC-2001 (first entry)
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.

XX Key Location/Qualifiers
FH Region 774..782
FT /note= "Antigenic epitope"

XX WO200168677-A2.

XX 20-SEP-2001.

XX 16-MAR-2001; 2001WO-US40328.

XX 16-MAR-2000; 2000US-0527487.

XX (GENZ) GENZYME CORP.

XX Nicolette CA;

XX WPI; 2001-616284/71.

XX N-PSDB; AAD19731.

XX Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties

XX
PS
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;

Query Match 98.9%; Score 6736; DB 22; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1243; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRHLYGQGVVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRHLYGQGVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNQVROVPLORLRIVRGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNQVROVPLORLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNQYIKANSKFIGITELQLRSLTEILKGGVLIORNPOLCYQDITLWKDIFHKNQLA 180
DB 121 DPLNNTPTVGTASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDITLWKDIFHKNQLA 180
QY 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
DB 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPPEGRTYFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPPEGRTYFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHMQEVTAEADGTORCEKSCPKARVCYGLGMEHLREVRVTSAN 360
DB 301 YNYLSTDVGSCTLVCPHMQEVTAEADGTORCEKSCPKARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVPFETLEEITGYLYISAWPDSL 420
DB 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVPFETLEEITGYLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLGSIWGLSLRELGSGLALIHNTLHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLGSIWGLSLRELGSGLALIHNTLHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVEGEGACHOLCARGHCWGPGTQCVCNCSOFLRGQEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDECVEGEGACHOLCARGHCWGPGTQCVCNCSOFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPFCVARC 600
DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPFCVARC 600
QY 601 PSGVKPDLSPYMWKFPDDEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSIVSAVVG 660
DB 601 PSGVKPDLSPYMWKFPDDEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILLIKRROQKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
DB 661 ILLVVVLGVVFGILLIKRROQKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKIILDEAYVMAGVGP 780
DB |||||
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKIILDEAYVMAGVGP 780
DB |||||
QY 781 YVSRLLGICLTSTVQLVTQMLPYCCLLDHVRENRRGRIGSQDLLNWCMIAGKMSYLEDVR 840
DB |||||
QY 781 YVSRLLGICLTSTVQLVTQMLPYCCLLDHVRENRRGRIGSQDLLNWCMIAGKMSYLEDVR 840
DB |||||
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPICKMALESILRRFT 900
DB |||||
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPICKMALESILRRFT 900
DB |||||
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKWM 960
DB |||||
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKWM 960
DB |||||
QY 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVDA 1020
DB |||||
QY 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVDA 1020
DB |||||
QY 1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
DB |||||
QY 1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
DB |||||
QY 1081 AGSDVFDGDLGMAAGKLSLPTHDPSLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
DB |||||
QY 1081 AGSDVFDGDLGMAAGKLSLPTHDPSLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
DB |||||
QY 1141 NQDVRPQPSREGPLPAARAGATLERAKTILSPGKGVVQKVFAGGAVENPEYLTPO 1200
DB |||||
QY 1141 NQDVRPQPSREGPLPAARAGATLERAKTILSPGKGVVQKVFAGGAVENPEYLTPO 1200
DB |||||
QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGDDVPV 1255
DB |||||
QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGDDVPV 1255
DB |||||

RESULT 3

AAB60167
ID AAB60167 standard; Protein; 1255 AA.

XX
AC AAB60167;

XX
DT 03-APR-2001 (first entry)

XX
DE HER2 transgene plasmid construct encoded protein.

XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.

XX
OS Homo sapiens.
OS Synthetic.

XX
PN W0200100244-A2.

XX
PD 04-JAN-2001.

XX
PF 23-JUN-2000; 2000WO-US17229.

XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.

XX
PA (GETH) GENENTECH INC.

XX
PI Erickeon S, Schwall R;

XX
DR WPI; 2001-061962/07.

XX
DR N-PSDB; AAF24297.

XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -

XX

PS Example 3; Fig 4; 92pp; English.

CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

XX
SQ Sequence 1255 AA;

Query Match 98.9%; Score 6736; DB 22; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1243; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASQVCTGDMKLRLPASPETHLDMLRHLVGGCQVQGNL 60
DB |||||
QY 1 MELAALCRWGLLALLPPGAASQVCTGDMKLRLPASPETHLDMLRHLVGGCQVQGNL 60
DB |||||
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVROVPLQRLRIVRGTOLEFEDYALAVLNG 120
DB |||||
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVROVPLQRLRIVRGTOLEFEDYALAVLNG 120
DB |||||
QY 121 DPLNNQYIKANSKIFIGITELQLRSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNQLA 180
DB |||||
QY 121 DPLNNQYIKANSKIFIGITELQLRSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNQLA 180
DB |||||
QY 181 LTLIDNRSRACHPCSPMKGSRCSWESSDQSLTRTVAGGCARCKGPLPTCCHEQC 240
DB |||||
QY 181 LTLIDNRSRACHPCSPMKGSRCSWESSDQSLTRTVAGGCARCKGPLPTCCHEQC 240
DB |||||
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPPEGRTYFGASCVTAC 300
DB |||||
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPPEGRTYFGASCVTAC 300
DB |||||
QY 301 YNYLSTDVGSCTLVCPLNQOEVTAEQTCRCKSKPCARVCYGLGHEHLREVRVTSAN 360
DB |||||
QY 301 YNYLSTDVGSCTLVCPLNQOEVTAEQTCRCKSKPCARVCYGLGHEHLREVRVTSAN 360
DB |||||
QY 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPSLP 420
DB |||||
QY 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPSLP 420
DB |||||
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGSLRLSRELGLALIHNTHLCFVHTV 480
DB |||||
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGSLRLSRELGLALIHNTHLCFVHTV 480
DB |||||
QY 481 PWDQLFRNPHOALLHTANRPEDECVGSEGLACHOLCARGHCWGPPTQCVNCSQFLRQEC 540
DB |||||
QY 481 PWDQLFRNPHOALLHTANRPEDECVGSEGLACHOLCARGHCWGPPTQCVNCSQFLRQEC 540
DB |||||
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
DB |||||
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
DB |||||
QY 601 PSGVKPDLSPYMPWKFPDEBACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660
DB |||||
QY 601 PSGVKPDLSPYMPWKFPDEBACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660
DB |||||
QY 661 ILLVVLGVVFGILIKRRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720
DB |||||
QY 661 ILLVVLGVVFGILIKRRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720
DB |||||
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKIILDEAYVMAGVGP 780
DB |||||
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKIILDEAYVMAGVGP 780
DB |||||
QY 781 YVSRLLGICLTSTVQLVTQMLPYCCLLDHVRENRRGRIGSQDLLNWCMIAGKMSYLEDVR 840
DB |||||
QY 781 YVSRLLGICLTSTVQLVTQMLPYCCLLDHVRENRRGRIGSQDLLNWCMIAGKMSYLEDVR 840
DB |||||
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPICKMALESILRRFT 900
DB |||||

Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHYHADGGKVPKWMALLESILRRFT 900
Qy 901 HQSDVMSYCVTWELMTGAKPYDGI PAREIPDLLEKGERLPDPPICTIDVYMINVKWM 960
Db 901 HQSDVMSYCVTWELMTGAKPYDGI PAREIPDLLEKGERLPDPPICTIDVYMINVKWM 960
Qy 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEGLPASPLDSTFYRSLLDDDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEGLPASPLDSTFYRSLLDDDDMDGLVDA 1020
Qy 1021 BEYLVPQGGFTCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEAEPRLAPSEG 1080
Db 1021 BEYLVPQGGFTCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEAEPRLAPSEG 1080
Qy 1081 AGSDVFDGLGMAKAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAKAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYLTPQ 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYLTPQ 1200
Qy 1201 GGAAPQHPHPPAFSPAFNLVYWDQPPERGAPPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFNLVYWDQPPERGAPPSTFKGTPTAENPEYGLDVPV 1255

RESULT 4

AAU74545
ID AAU74545 standard; Protein; 1255 AA.
AC AAU74545;
XX
XX
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.

Human; ErbB; epidermal growth factor receptor; receptor;
anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
stomach; endometrium, salivary gland; lung; kidney; colon; colorectum;
thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
glioma disorder; astrocytoma disorder; hypothalamic disorder;
glandular disorder; macrophagal disorder; epithelial disorder;
stromal disorder; blastocoeleic disorder; inflammatory disorder;
angiogenic disorder; immunological disorder.

XX Homo sapiens.
OS
XX
XX US2002001587-A1.
XX
XX 03-JAN-2002.
XX
XX 16-MAR-2001; 2001US-0811123.
XX
XX 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
XX (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX
XX Erickson S, Schwall R, Sliwkowski M;
PI
XX WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
XX Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
XX Example 3; Fig 7; 93pp; English.
PS
XX
XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytic, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoeleic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.

XX Sequence 1255 AA;

Query Match 98.9%; Score 6736; DB 23; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1243; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPTHLDMLRHLQSCQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPTHLDMLRHLQSCQVQGNL 60
Qy 61 ELTYLPTNASLFLQDIQEVQGYVLIHNVQVPLQRLRIVRGTLQEDNYVALVDNG 120
Db 61 ELTYLPTNASLFLQDIQEVQGYVLIHNVQVPLQRLRIVRGTLQEDNYVALVDNG 120
Qy 121 DPLNNQYIKANSKFITELQLRSITLTKGGVLIQNPOLQCYQDTILWKFDFHNNQLA 180
Db 121 DPLNNQYIKANSKFITELQLRSITLTKGGVLIQNPOLQCYQDTILWKFDFHNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVAGGCAKRGPLTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVAGGCAKRGPLTDCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 300
Qy 301 YNYLSTDVSGCTLVCPHNLNQVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGCTLVCPHNLNQVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IOEPAGCKKIFGSLAFLESFDPGDPASNTAPLOPEQLQVFTLEBITGYLISAMPDSL 420
Db 361 IOEPAGCKKIFGSLAFLESFDPGDPASNTAPLOPEQLQVFTLEBITGYLISAMPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLRELGSLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLRELGSLALIHNNTHLCFVHTV 480
Qy 481 PWDQLFRNPHOALLHTANRDECEVGEGLACHQJARGHCWGPGTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHOALLHTANRDECEVGEGLACHQJARGHCWGPGTQCVCNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEPEADQCACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEPEADQCACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCHTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCHTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRIKTEL 720
Db 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRIKTEL 720
Qy 721 RKVKVLGSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICITSTVQLVTQLMPYGCLLDHRNENRGLGSQDLNLCWQIAKMSYLEDVR 840

Db 781 YVSRLLGICLTSTVQLVTLMPYGCLLDHVRENRGLSGQDLLNMCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNMALESILRRRT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNMALESILRRRT 900
Qy 901 HOSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMWKCMW 960
Db 901 HOSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMWKCMW 960
Qy 961 IDSECRPRELVSEFSRMRDPQRFVWJONEDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
Db 961 IDSECRPRELVSEFSRMRDPQRFVWJONEDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
Qy 1021 ESYLVPOQGFCCPDAPAGAGGVMVHRSSSTRSGGDLTLGLEPSEEBAPRPLAPSG 1080
Db 1021 ESYLVPOQGFCCPDAPAGAGGVMVHRSSSTRSGGDLTLGLEPSEEBAPRPLAPSG 1080
Qy 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVDVFAFGGAVENPEYLTQ 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVDVFAFGGAVENPEYLTQ 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 5

AAW01111
ID AAW01111 standard; Protein; 1255 AA.
AC AAW01111;
DT 01-JAN-1997 (first entry)
XX HER-2/neu protein.
DE
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "claimed domain, useful for immunisation"
XX
PN WO9630514-A1.
XX
PD 03-OCT-1996.
XX
PF 28-MAR-1996; 96WO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNITW) UNIV WASHINGTON.
XX
PI Cheever MA, Disis MB;
XX
DR WPI: 1996-455361/45.
DR N-PSDB; AAT40739.
XX
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 98.8%; Score 6730; DB 17; Length 1255;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRPLASPETHLDMRLHLVYQGCQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRPLASPETHLDMRLHLVYQGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQYVLI AHNQVRQVPLQRLRI VRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQYVLI AHNQVRQVPLQRLRI VRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNQYIKANSKFPITELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGSRCSGSESDCQSLTRTV CAGGCARCKGPLPTDCHEOC 240
Db 181 LTLIDTNRSRACHPCSPMKSGSRCSGSESDCQSLTRTV CAGGCARCKGPLPTDCHEOC 240
Qy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNLSLTDVGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNLSLTDVGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFLEETITGVLYISAMPDLSLP 420
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFLEETITGVLYISAMPDLSLP 420
Qy 421 DLSVPQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSLALIHNNHLCFVHTV 480
Db 421 DLSVPQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSLALIHNNHLCFVHTV 480
Qy 481 PWDQLFRNPHOALLHTANRPEDECYEGGLACHOLCARGHCWPGPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECYEGGLACHOLCARGHCWPGPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLOGLPREYVYNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
Db 541 VEECRVLOGLPREYVYNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIYSAVVG 660
Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIYSAVVG 660
Qy 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTLMPYGCLLDHVRENRGLSGQDLLNMCQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLMPYGCLLDHVRENRGLSGQDLLNMCQIAKMSYLEDVR 840

QY 841 LVHRDLAARNVLKSPBNHVKITDFGLARLLDIDETEHADGGKVP1KWMALLESILRRFT 900
 DB 841 LVHRDLAARNVLKSPBNHVKITDFGLARLLDIDETEHADGGKVP1KWMALLESILRRFT 900
 QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
 DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
 QY 961 IDSECRPRFRELVSFGRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGDLVDA 1020
 DB 961 IDSECRPRFRELVSFGRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGDLVDA 1020
 QY 1021 EYLVPQGGFCPPAPCAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 DB 1021 EYLVPQGGFCPPAPCAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 QY 1081 AGSDVFDGDLGMAAGKGLQSLTPHDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 DB 1081 AGSDVFDGDLGMAAGKGLQSLTPHDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1141 NQPDVVPQPSPREGPIPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYILTPQ 1200
 DB 1141 NQPDVVPQPSPREGPIPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYILTPQ 1200
 QY 1201 GGAAPQHPPPAFSPFDNLYYWDQDPPERCAPPSTFKGTPTAENPEYILGDDVPV 1255
 DB 1201 GGAAPQHPPPAFSPFDNLYYWDQDPPERCAPPSTFKGTPTAENPEYILGDDVPV 1255

RESULT 6

AAW92406
 ID AAW92406 standard; Protein: 1255 AA.

XX AAW92406;

XX 21-APR-1999 (first entry)

XX Human HER-2/neu oncogene protein.

XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 KW malignancy; treatment; tumour.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT Region 676..1255

FT /note= "region which elicits immune response"

XX US5869445-A.

XX 09-FEB-1999.

XX 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.

PR 17-MAR-1993; 93US-0033644.

PR 12-AUG-1993; 93US-0106112.

PR 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

DR WPI; 1999-152835/13.

DR N-PSDB; AAX01912.

XX Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours

XX Claim 3; Column 31-38; 26pp; English.

PS This sequence represents the human HER-2/neu oncogene protein. A fragment

XX

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.

XX Sequence 1255 AA;

Query Match 98.8%; Score 6730; DB 20; Length 1255;
 Beat Local Similarity 98.9%; Pred. No. 0;
 Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCQVVOGNL 60
 DB 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCQVVOGNL 60
 QY 61 ELTYLPTNASLSFLQDIQEVQYVLIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
 DB 61 ELTYLPTNASLSFLQDIQEVQYVLIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
 QY 121 DPLNQVIKANSKFIGITELQLRSLTEILKGVLIQRNPOLCYODTILWKDI FHKNNOLA 180
 DB 121 DPLNQVIKANSKFIGITELQLRSLTEILKGVLIQRNPOLCYODTILWKDI FHKNNOLA 180
 QY 181 LTLIDTNRSRACHPCSPMCKGSRGWCSESDCQSLTRTVAGGCARCKGPLTDCCHQC 240
 DB 181 LTLIDTNRSRACHPCSPMCKGSRGWCSESDCQSLTRTVAGGCARCKGPLTDCCHQC 240
 QY 241 AAGCTGPKHSDCLACHLHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACHLHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSLVCPPLHNOBVTAEADGTORCEKSKPCARVCYGLGMEHLREVAVTSAN 360
 DB 301 YNYLSTDVGSLVCPPLHNOBVTAEADGTORCEKSKPCARVCYGLGMEHLREVAVTSAN 360
 QY 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVPFETLEETGLYISAWPDSL 420
 DB 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVPFETLEETGLYISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLOGIGISWGLRSLRELGLALIHNNTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTLOGIGISWGLRSLRELGLALIHNNTHLCFVHTV 480
 QY 481 PWDLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
 DB 481 PWDLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
 QY 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
 DB 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
 QY 601 PSGVKPDLSPYMPWKFPDEEGACQPCPNCTHSCVDLDDKGCAPAEQASPLTSTVSAVG 660
 DB 601 PSGVKPDLSPYMPWKFPDEEGACQPCPNCTHSCVDLDDKGCAPAEQASPLTSTVSAVG 660
 QY 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOSTELVEPLTPSCAMPNQAMRLKETEL 720
 DB 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOSTELVEPLTPSCAMPNQAMRLKETEL 720
 QY 721 RKVVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILD EAYVMAGVSP 780
 DB 721 RKVVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILD EAYVMAGVSP 780
 QY 781 YVSRLLIGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQDLNWCNQIAKGSYLEDDR 840
 DB 781 YVSRLLIGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQDLNWCNQIAKGSYLEDDR 840
 QY 841 LVHRDLAARNVLKSPBNHVKITDFGLARLLDIDETEHADGGKVP1KWMALLESILRRFT 900
 DB 841 LVHRDLAARNVLKSPBNHVKITDFGLARLLDIDETEHADGGKVP1KWMALLESILRRFT 900

Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
Dy 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
Qy 961 IDSECRPRFRELVSFMSHARDPQRFVVIQNEGLGPASPLDSTYRSILLEDDMGDLVDA 1020
Dy 961 IDSECRPRFRELVSFMSHARDPQRFVVIQNEGLGPASPLDSTYRSILLEDDMGDLVDA 1020
Qy 1021 EYLVLPOOGFFCPDAPAGAGMWHHRHSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Dy 1021 EYLVLPOOGFFCPDAPAGAGMWHHRHSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Dy 1081 AGSDVFDGDLGMAAGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPOPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Dy 1141 NOPDVRPOPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
Dy 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 7
ID AAB21198 standard; protein; 1255 AA.
AC AAB21198;
XX 12-JAN-2001 (first entry)
XX Human HER-2/neu protein.
XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
XX breast cancer; prostate cancer; ovarian cancer; lung cancer;
XX colon cancer.
XX Homo sapiens.
XX W0200044999-A1.
XX 03-AUG-2000.
XX 28-JAN-2000; 2000WO-US02164.
XX 29-JAN-1999; 99US-0117976.
XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHKLINE BEECHAM.
XX Cheever MA, Gheysen D;
XX WPI; 2000-505976/45.
XX N-PSDB; AAA89736.
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
XX useful for vaccinating against breast, ovarian, colon, lung and
XX prostate cancers -
XX Claim 52; Fig 7; 128pp; English.
XX The present sequence is the human HER-2/neu protein. It is a member of
XX the tyrosine kinase family of receptor-like glycoproteins and shows
XX homology to the epidermal growth factor receptor (EGFR). It probably
XX plays a part in cell growth and/or differentiation. The HER-2/neu
XX gene is an oncogene. An HER-2/neu fusion protein comprising a
XX HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
XX domain may be used to treat or prevent cancer by eliciting or
XX enhancing an immune response to the HER-2/neu protein. It may be used
XX to treat malignancies such as breast, ovarian, colon, lung and
XX prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 98.8%; Score 6730; DB 21; Length 1255;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDKMLRLPASPTHLDMLRHLYGCCQVQGNL 60
Dy 1 MELAALCRWGLLLALLPPGAASQVCTGTDKMLRLPASPTHLDMLRHLYGCCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQYVLI AHNVROVPLQRLRIVRGRTQLPEDNYALAVLDNG 120
Dy 61 ELTYLPTNASLSFLQDIQEVQYVLI AHNVROVPLQRLRIVRGRTQLPEDNYALAVLDNG 120
Qy 121 DPLNNYIKANSKFIGITELQLRSLTEILKGGVLIQRPNQLCYQDTILWKDIFHNQOLA 180
Dy 121 DPLNNYIKANSKFIGITELQLRSLTEILKGGVLIQRPNQLCYQDTILWKDIFHNQOLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
Dy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Dy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHNOEVTAEQDQRCCKSPCARVYCVGLGMEHLREVRAVTSAN 360
Dy 301 YNYLSTDVGSCTLVCPLNHNOEVTAEQDQRCCKSPCARVYCVGLGMEHLREVRAVTSAN 360
Qy 361 IQBFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAMPDPL 420
Dy 361 IQBFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAMPDPL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGRLSRLRELGLSLALIHNNHLCFVHTV 480
Dy 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGRLSRLRELGLSLALIHNNHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWPGPTQCVNCSQFLRGOEC 540
Dy 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWPGPTQCVNCSQFLRGOEC 540
Qy 541 VEESRVLQGLPREVYNARHCLPCHPECPONGSVTCFGEADQCVACHYKDPFCVCARC 600
Dy 541 VEESRVLQGLPREVYNARHCLPCHPECPONGSVTCFGEADQCVACHYKDPFCVCARC 600
Qy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSTVSAVVG 660
Dy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSTVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720
Dy 661 ILLVVLGVVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
Dy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
Qy 781 YVSRLGICLTSTVQLVLTOLMPYGCLLDHVRENHGRGLSODLLNWCQIAKMSYLEDVR 840
Dy 781 YVSRLGICLTSTVQLVLTOLMPYGCLLDHVRENHGRGLSODLLNWCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESILRRFT 900
Dy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESILRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
Dy 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
Qy 961 IDSECRPRFRELVSFMSHARDPQRFVVIQNEGLGPASPLDSTYRSILLEDDMGDLVDA 1020

Db 961 IDSECRPRELVSEFSRMARDPQREWVQINEDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020
QY 1021 EYLVVPOQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVVPOQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSESDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSESDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVVRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVVRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPOPHPPAFSPAFDNLVYWDQPPERPAGPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPOPHPPAFSPAFDNLVYWDQPPERPAGPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 8

AA584780
ID AAY84780 standard; Protein; 1255 AA.
XX AAY84780;
AC AAY84780;
DT 08-AUG-2000 (first entry)
XX Amino acid sequence of the SPLICE erbb-2 receptor protein.
DE SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.

Homo sapiens.

XX WO200020579-A1.

XX 13-APR-2000.

XX 01-OCT-1999; 99WO-CA00912.

XX 02-OCT-1998; 98US-0165192.

XX (UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

XX WPI; 2000-303768/26.

XX N-PSDB; AAA14812.

PT Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
PT erbb-2, inhibitors of the protein are useful for treatment of cancer -

PS Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbb-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbb-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbb-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbb-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbb-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.

XX Sequence 1255 AA;

| Query Match | 98.8% | Score 6730; | DB 21; | Length 1255; |
|-----------------------|--------------------|---|------------|--------------|
| Best Local Similarity | 98.9% | Pred. No. 0; | | |
| Matches 1241; | Conservative | 3; | Mismatches | 11; |
| | | | Indels | 0; |
| | | | Gaps | 0; |
| QY 1 | MELAAALCRGGLLALLLP | PGAASTQVCTGTDMLKRLPASPTHLDMLRHLVQGCQVQGNL | 60 | |
| Db 1 | MELAAALCRGGLLALLLP | PGAASTQVCTGTDMLKRLPASPTHLDMLRHLVQGCQVQGNL | 60 | |
| QY 61 | ELTYLPTNASLFLQDIQ | EVQGVVLIHAHQVROVPLQRLIRVGTOLFEONYALAVLDNG | 120 | |
| Db 61 | ELTYLPTNASLFLQDIQ | EVQGVVLIHAHQVROVPLQRLIRVGTOLFEONYALAVLDNG | 120 | |
| QY 121 | DPLNNQYIKANSKFIGIT | ELQLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA | 180 | |
| Db 121 | DPLNNTPVTGASPGGL | RELQSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA | 180 | |
| QY 181 | LTLIDNRSRACHPCSP | CMCKSGRCWGESSEDCOSLTRTVCCAGGCARCKGPLTDCCHQEC | 240 | |
| Db 181 | LTLIDNRSRACHPCSP | CMCKSGRCWGESSEDCOSLTRTVCCAGGCARCKGPLTDCCHQEC | 240 | |
| QY 241 | AAGCTGPKHSDCLACL | HFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP | 300 | |
| Db 241 | AAGCTGPKHSDCLACL | HFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP | 300 | |
| QY 301 | YNYLSTDVGSCTLV | CPHLNQEVTAEDGTQCEKCKPCARVCYGLGMEHLREVRAVTSAN | 360 | |
| Db 301 | YNYLSTDVGSCTLV | CPHLNQEVTAEDGTQCEKCKPCARVCYGLGMEHLREVRAVTSAN | 360 | |
| QY 361 | IQEFAGCKIKFSLAFL | PESFDGDPASNTAPLOPEOLQVPETLEEITGYLIYISAWPDSLP | 420 | |
| Db 361 | IQEFAGCKIKFSLAFL | PESFDGDPASNTAPLOPEOLQVPETLEEITGYLIYISAWPDSLP | 420 | |
| QY 421 | DLVSFONLQVIRGRIL | HNCAYSITLQGLGISWGLRSLRELGSGLALIHHTHLCFVHTV | 480 | |
| Db 421 | DLVSFONLQVIRGRIL | HNCAYSITLQGLGISWGLRSLRELGSGLALIHHTHLCFVHTV | 480 | |
| QY 481 | PWDQLFRNPHQALLHT | ANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSFRLQOEC | 540 | |
| Db 481 | PWDQLFRNPHQALLHT | ANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSFRLQOEC | 540 | |
| QY 541 | VVECRVLQGLPREYV | NARHCLPCHPECQONGSVTCFGEADQCVACAHYKDPFPCVARC | 600 | |
| Db 541 | VVECRVLQGLPREYV | NARHCLPCHPECQONGSVTCFGEADQCVACAHYKDPFPCVARC | 600 | |
| QY 601 | PSGVKPDLSYMPIWK | FPDEEGACQPCINCTHSCVDLDDKCPA2ORASPLTISVSAVG | 660 | |
| Db 601 | PSGVKPDLSYMPIWK | FPDEEGACQPCINCTHSCVDLDDKCPA2ORASPLTISVSAVG | 660 | |
| QY 661 | ILLVVLGVVFGILIK | RROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL | 720 | |
| Db 661 | ILLVVLGVVFGILIK | RROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL | 720 | |
| QY 721 | RKVVLGSGAGFTVYK | GIWIPDGENVKIPVAIKVRENTSPKANKIILDEAYVMAGVGP | 780 | |
| Db 721 | RKVVLGSGAGFTVYK | GIWIPDGENVKIPVAIKVRENTSPKANKIILDEAYVMAGVGP | 780 | |
| QY 781 | VYSRLGICLTSTVOL | VTOLMPYGCILLDHVRENRLGSGQDILLNWCQIAGKHSYLEDVR | 840 | |
| Db 781 | VYSRLGICLTSTVOL | VTOLMPYGCILLDHVRENRLGSGQDILLNWCQIAGKHSYLEDVR | 840 | |
| QY 841 | LVRDLAARNVLKSPN | HVKITDFGLARLLDIDETEHADGGKVPIKNWALESLRRRFT | 900 | |
| Db 841 | LVRDLAARNVLKSPN | HVKITDFGLARLLDIDETEHADGGKVPIKNWALESLRRRFT | 900 | |
| QY 901 | HQSDVMSYGVTVWEL | MTFGAKPYDGIIPAREIPDLLEKGERLPPQPICTIDVYIMVKCM | 960 | |
| Db 901 | HQSDVMSYGVTVWEL | MTFGAKPYDGIIPAREIPDLLEKGERLPPQPICTIDVYIMVKCM | 960 | |
| QY 961 | IDSECRPRELVSEFS | RMARDPQRFVQINEDLGPASPLDSTFYRSLLEDDDDMGDLVDA | 1020 | |
| Db 961 | IDSECRPRELVSEFS | RMARDPQRFVQINEDLGPASPLDSTFYRSLLEDDDDMGDLVDA | 1020 | |
| QY 1021 | EYLVVPOQGGFFCP | DPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG | 1080 | |

Db 1021 EYLVFQOQFPCDPAPGAGMVHHRSSSTRSGGDLTLGLERPEEEAPRPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAKGQLSLPHDPSPLQRYSEDPTVLPSPETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAKGQLSLPHDPSPLQRYSEDPTVLPSPETDGVVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPPPAPSPAFDNLYYWDQPPERCAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPPPAPSPAFDNLYYWDQPPERCAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 9
AA885458
ID AA885458 standard; Protein; 1255 AA.
XX
AC AA885458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
PN WO200153463-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
XX (CORI-) CORIXA CORP.
XX
XX Cheever MA, Hand-Zimmermann S;
XX
XX WPI; 2001-476112/51.
DR N-PSDB; AAH23392.
XX
XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -
XX
XX Claim 2; Page 41-46; 49pp; English.
XX
XX The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;
Query Match 98.8%; Score 6730; DB 22; Length 1255;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
Qy 1 MELAAALCRWGLLLALLLPPGAASSTVCTGTDMLRLPASPEHLDMLRHLYYQGCVVQGNL 60
Db 1 MELAAALCRWGLLLALLLPPGAASSTVCTGTDMLRLPASPEHLDMLRHLYYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLODIEVQGVVLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 1200

Db 61 ELTYLPTNASLSFLODIEVQGVVLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNQYIKANSKFIGITELQRLSLTEILKGGVLIQORNPOLCYODTTLWKDI FHKNNOLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTTLWKDI FHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKCKGSRGSESSDCQSLTRTVCCAGGCARCKGPLPDDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKCKGSRGSESSDCQSLTRTVCCAGGCARCKGPLPDDCCHEQC 240
Qy 241 AGCTGPKGSDCLACLHFNHSGICELHCPALVYNTDTPFESMPNPEGRYTFGASCVTACP 300
Db 241 AGCTGPKGSDCLACLHFNHSGICELHCPALVYNTDTPFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHQNVEADGTORCEKSKPCARVCYGLGMEHLREVRVAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNHQNVEADGTORCEKSKPCARVCYGLGMEHLREVRVAVTSAN 360
Qy 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEEITGYLYISAWPDSL 420
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEEITGYLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGISWGLSLRELGLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGISWGLSLRELGLALIHNTHLFCFVHTV 480
Qy 481 PWDOLFRNPHQALLHTANRPEDECVEGLACHOLCARGHCWGPGTQCVCNCSOFLRQEC 540
Db 481 PWDOLFRNPHQALLHTANRPEDECVEGLACHOLCARGHCWGPGTQCVCNCSOFLRQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFFCVARC 600
Qy 601 PSGVKPDLSPYPIWKFPDEEGACQPCINCTHSCVDLDDKGCPEAQASPLTSISAVVG 660
Db 601 PSGVKPDLSPYPIWKFPDEEGACQPCINCTHSCVDLDDKGCPEAQASPLTSISAVVG 660
Qy 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLGSDLLNWCNQIAKGMYSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLGSDLLNWCNQIAKGMYSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMMALESILRRRFT 900
Qy 901 HQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKCM 960
Db 901 HQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKCM 960
Qy 961 IDSECRPRFRELVSFERSMARDPQRFVJONEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Db 961 IDSECRPRFRELVSFERSMARDPQRFVJONEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Qy 1021 EYLVFQOQFPCDPAPGAGMVHHRSSSTRSGGDLTLGLERPEEEAPRPLAPSEG 1080
Db 1021 EYLVFQOQFPCDPAPGAGMVHHRSSSTRSGGDLTLGLERPEEEAPRPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAKGQLSLPHDPSPLQRYSEDPTVLPSPETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAKGQLSLPHDPSPLQRYSEDPTVLPSPETDGVVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVFAFGGAVENPEYLTPO 1200

Db 1141 NQDVRQPPSPREGPLPAARPAAGATLERPKTSLSPGKNGVVKVAFGGAIVENPEYLTQP 1200
AAG88267
QY 1201 GGAAPHPHPAFSPAFDNLYYDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
DB 1201 GGAAPHPHPAFSPAFDNLYYDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
OS Homo sapiens.
XX
XX WO200141787-A1.
XX
PD 14-JUN-2001.
XX
XX 11-DEC-2000; 2000MO-US33591.
XX
XX 10-DEC-1999; 99US-0458299.
XX
XX (EPIM-) EPIMMUNE INC.
XX
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
XX WPI; 2001-374995/39.
XX
XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX
XX Disclosure; Page 15; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability, and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
XX Sequence 1255 AA;

Query Match 98.8%; Score 6730; DB 22; Length 1255;

Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MELAAALCRWGLLALLPPGAASSTOVCTGTDMKURLPAS PETHLDMLRHLVYGCQVVOGNL 60
DB 1 MELAAALCRWGLLALLPPGAASSTOVCTGTDMKURLPAS PETHLDMLRHLVYGCQVVOGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLQRLRIVRGTOFLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLQRLRIVRGTOFLFEDNYALAVLDNG 120
QY 121 DPLNNQYIKANSKFIGITELQRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
QY 181 LTLIDTNRSPACHPCSPMKGSCWGESSEDCSLTRTV CAGGCARCKGPLPTCCHEQC 240
DB 181 LTLIDTNRSPACHPCSPMKGSCWGESSEDCSLTRTV CAGGCARCKGPLPTCCHEQC 240
QY 241 AAGCTGPKGSDCLACLHFNHSGICELHCPALVYNTDT FESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKGSDCLACLHFNHSGICELHCPALVYNTDT FESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSP CARVCYGLGMEHLREVRVTSAN 360
DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSP CARVCYGLGMEHLREVRVTSAN 360
QY 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVP ETLEEITGYLYISAWPDSLP 420
DB 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVP ETLEEITGYLYISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRIHLHGAYSILTLOGLGISWLGSLRLSRLGSLALIHNTLHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRIHLHGAYSILTLOGLGISWLGSLRLSRLGSLALIHNTLHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDCEVGEGLACHOLCARGH CWPBGPTQCVNCSQFLRGQEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDCEVGEGLACHOLCARGH CWPBGPTQCVNCSQFLRGQEC 540
QY 541 VEBCRVLQGLPREVYNARHCLPCHPECPONGSVTCFGE PNDQCVACAHYKDPFCVARC 600
DB 541 VEBCRVLQGLPREVYNARHCLPCHPECPONGSVTCFGE PNDQCVACAHYKDPFCVARC 600
QY 601 PSQVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDK GCPAEQASPLTISVAVVG 660
DB 601 PSQVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDK GCPAEQASPLTISVAVVG 660
QY 661 ILLVVLGVVFGILIKRQOKIRKYMRRLLQETELVEPLTP SGAMPNQAMRILKETEL 720
DB 661 ILLVVLGVVFGILIKRQOKIRKYMRRLLQETELVEPLTP SGAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTS PKANKETILDEAYVMAGVSP 780
DB 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTS PKANKETILDEAYVMAGVSP 780
QY 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQD LLDNWCQIAKMSYLEVDV 840
DB 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQD LLDNWCQIAKMSYLEVDV 840
QY 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLIDIDETEHAD GGVKPIKMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLIDIDETEHAD GGVKPIKMALESILRRFT 900
QY 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLEKGERL POPPITIDVYMIWVKWM 960
DB 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLEKGERL POPPITIDVYMIWVKWM 960
QY 961 IDSECRPRFRELVSFERNARDPQRFVVIQNEDLGPASPLD STFYRSLLEDDDMGLVDA 1020
DB 961 IDSECRPRFRELVSFERNARDPQRFVVIQNEDLGPASPLD STFYRSLLEDDDMGLVDA 1020
QY 1021 EYLVFPQQGFFCDDPAPGAGGMVHRRHSSTSGGGDLTLGL ESEEEAPRSLAPSEG 1080
DB 1021 EYLVFPQQGFFCDDPAPGAGGMVHRRHSSTSGGGDLTLGL ESEEEAPRSLAPSEG 1080


```

Db 1021 EEYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGKGLQSLPTHDPSPQLQRYSDPTVPLPSETDGVYAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGKGLQSLPTHDPSPQLQRYSDPTVPLPSETDGVYAPLTCSPQPEYV 1140
Qy 1141 NOPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVDVFAFGGAVENPEYLTQP 1200
Db 1141 NOPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVDVFAFGGAVENPEYLTQP 1200
Qy 1201 GGAAPQHPPPPAFSPADNLYWDDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPPPPAFSPADNLYWDDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 11
AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX
AC AAE24067;
DT 23-SEP-2002 (first entry)
DE
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX
OS Homo sapiens.
XX
PN WO200222636-A1.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowser LM;
XX
DR WPI; 2002-471192/50.
XX
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;

Query Match 98.8%; Score 6730; DB 23; Length 1255;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASSTVCTGCTDMKRLPASPTHLDMLRHLVGGCQVOGNL 60
Db 1 MELAALCRWGLLALLPPGAASSTVCTGCTDMKRLPASPTHLDMLRHLVGGCQVOGNL 60

```

```

Qy 61 ELTYLPTNASLFLQDIOEQVYVLI AHNOVROVPLQRLRIVRGCTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIOEQVYVLI AHNOVROVPLQRLRIVRGCTQLFEDNYALAVLDNG 120
Qy 121 DPLNNQYIKANSKFIGITELQLRSLTEILKGGVLIQRNPQLCYQDITLWKDIFHKNNOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDITLWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPCKGSRGWGSESDCQSLTRTVGAGGCARCKGPLPTCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPCKGSRGWGSESDCQSLTRTVGAGGCARCKGPLPTCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREAVRATSAN 360
Db 301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREAVRATSAN 360
Qy 361 IQBFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEETGTYLISAMPDSL 420
Db 361 IQBFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEETGTYLISAMPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGRLSRLRELGLSLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGRLSRLRELGLSLALIHNTLHLCFVHTV 480
Qy 481 PWQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
Db 481 PWQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPKWKFPDEGACQPCPINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660
Db 601 PSGVKPDLSPYMPKWKFPDEGACQPCPINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVWAGVSP 780
Db 721 RKVKVLSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVWAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLMPYGCCLLDHVRENRLGSDQLLNWCQIAGKMSYLEOVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLMPYGCCLLDHVRENRLGSDQLLNWCQIAGKMSYLEOVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWKMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWKMALESILRRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDYVMIMVKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDYVMIMVKWM 960
Qy 961 IDSECRPRFRELVESEFSRMDARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDGLVDA 1020
Db 961 IDSECRPRFRELVESEFSRMDARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDGLVDA 1020
Qy 1021 EYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Db 1021 EYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGKGLQSLPTHDPSPQLQRYSDPTVPLPSETDGVYAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGKGLQSLPTHDPSPQLQRYSDPTVPLPSETDGVYAPLTCSPQPEYV 1140
Qy 1141 NOPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVDVFAFGGAVENPEYLTQP 1200

```


Db 1021 EEVLVPOQGFCCPDAPAGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 QY 1081 AGSDVFDGLGMGAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGLGMGAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVKDVFAGGAVENPEYLTPO 1200
 Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVKDVFAGGAVENPEYLTPO 1200
 QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYGLDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 13

AAM51143

ID AAM51143 standard; Protein; 1255 AA.

XX AC

XX AAM51143;

XX 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

XX Key

XX Domain

XX Location/Qualifiers

XX 1..653

XX /note= "extracellular domain"

XX Domain

XX 676..1255

XX /note= "intracellular domain"

XX Domain

XX 990..1255

XX /note= "phosphorylation domain"

XX WO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

XX or enhancing an immune response to the protein, has Her-2/neu

XX extracellular domain fused to Her-2/neu intracellular or

XX phosphorylation domain -

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein

XX or c-erbB2), an oncogenic self-protein and target for anti-cancer

XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed

XX in a variety of cancers, including breast, ovarian, colon, lung and

XX prostate cancer. Her-2/neu is a member of the tyrosine kinase

XX family of receptor-like glycoproteins. It comprises an extracellular

XX domain with homology to the epidermal growth factor receptor

XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal

XX intracellular domain that also shows homology to EGFR. Its

XX overexpression correlates with a poor prognosis in breast and

XX ovarian cancers. The invention provides Her-2/neu fusion

XX proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its beltapd fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC transfecting cells of an animal ex vivo with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.

XX SQ Sequence 1255 AA;

Query Match 98.8%; Score 6730; DB 23; Length 1255;
 Best Local Similarity 98.9%; Pred. No. 0;
 Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKLRLPASPETHLDMRLHYQGCQVQGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKLRLPASPETHLDMRLHYQGCQVQGNL 60
 QY 61 ELTYLPTNASLFLQDIOEQVGYLIAHNOVQVPLQRLRIVRGTOFLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLFLQDIOEQVGYLIAHNOVQVPLQRLRIVRGTOFLFEDNYALAVLDNG 120
 QY 121 DPLNNQYIKANSKFIGITELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
 Db 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
 QY 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCOSLTRTVAGGCARCKGPLPTCCHEQC 240
 Db 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCOSLTRTVAGGCARCKGPLPTCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
 Db 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
 QY 361 IQEFAGCKITFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 420
 Db 361 IQEFAGCKITFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 420
 QY 421 DLSVFONLQVIRGRIILHNGAYSILTLOGLGISWLGRLSLRELGSGLALIHNNTHLCFVHTV 480
 Db 421 DLSVFONLQVIRGRIILHNGAYSILTLOGLGISWLGRLSLRELGSGLALIHNNTHLCFVHTV 480
 QY 481 PWDLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTOCVNCQSFRLQGE 540
 Db 481 PWDLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTOCVNCQSFRLQGE 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFPCVARC 600
 Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFPCVARC 600
 QY 601 PSQVKPDLSPYMPITWKEPDEBACQPCPINCTHSCVDLDDKGCAPAEORASPLTSTVSAVVG 660
 Db 601 PSQVKPDLSPYMPITWKEPDEBACQPCPINCTHSCVDLDDKGCAPAEORASPLTSTVSAVVG 660
 QY 661 ILLVVLGVVVGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMNQAMRILKETEL 720
 Db 661 ILLVVLGVVVGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMNQAMRILKETEL 720
 QY 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
 Db 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780

781 YVSRLLGICLTSTVQLVLTQMPYGCCLLDHVRENRGRGLSGDQLLNWCQIAKGNLSYLEDVR 840
|
781 YVSRLLGICLTSTVQLVLTQMPYGCCLLDHVRENRGRGLSGDQLLNWCQIAKGNLSYLEDVR 840
|
841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIAKGNLSYLEDVR 900
|
841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIAKGNLSYLEDVR 900
|
901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYIMVKCWM 960
|
901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYIMVKCWM 960
|
961 IDSECRPRFRELVSERWARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLDVA 1020
|
961 IDSECRPRFRELVSERWARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLDVA 1020
|
1021 EYLVPQOQFFCDDPAPGAGMWHHRSSSTRSGGDLTLGLERPESEEAAPRSLPASEG 1080
|
1021 EYLVPQOQFFCDDPAPGAGMWHHRSSSTRSGGDLTLGLERPESEEAAPRSLPASEG 1080
|
1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
|
1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
|
1141 NQPDVREPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
|
1141 NQPDVREPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
|
1201 GGAAPOHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
|
1201 GGAAPOHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
|

RESULT 14

AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
XX
DT 05-JUN-2002 (first entry)
XX
XX Human Her-2/neu polypeptide.
DE
XX Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
XX WO200213847-A2.
XX
XX
PD 21-FEB-2002.
XX
XX 13-AUG-2001; 2001WO-US25408.
XX
XX 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
PI
XX WPI: 2002-280741/32.
XX
DR N-PSDB; ABK10730.
DR
XX
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.8%; Score 6730; DB 23; Length 1255;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MELAALCRGGLLLALLPPGAASTQVCTGDMKURLPASPEETHLDMLRHLYQGQVVOGNL 60
|
DB 1 MELAALCRGGLLLALLPPGAASTQVCTGDMKURLPASPEETHLDMLRHLYQGQVVOGNL 60
|
QY 61 ELYLPTNASLSFLQDIQEVQGVVLIARNOVROVPLQRLIRVRGTOLFEDNYALAVLDNG 120
|
DB 61 ELYLPTNASLSFLQDIQEVQGVVLIARNOVROVPLQRLIRVRGTOLFEDNYALAVLDNG 120
|
QY 121 DPLNNQYIKANSKFIGITELQLRSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
|
DB 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
|
QY 181 LTLIDTNRGRACHPCSPCKGRCWGESSEDCQSLTRTVCCAGCARCKGPLTDCCHQEC 240
|
DB 181 LTLIDTNRGRACHPCSPCKGRCWGESSEDCQSLTRTVCCAGCARCKGPLTDCCHQEC 240
|
QY 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
|
DB 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
|
QY 301 YNYLSTDVGSCTLVCPHMQEVTAEQGTORCEKSKPCARVCYGLGMEHUREVRAVTSAN 360
|
DB 301 YNYLSTDVGSCTLVCPHMQEVTAEQGTORCEKSKPCARVCYGLGMEHUREVRAVTSAN 360
|
QY 361 IOEFAGCKKIFGSLAFLESFDDGPASNTAPLOEQVFEETLEETIGYLYISAWPDSLP 420
|
DB 361 IOEFAGCKKIFGSLAFLESFDDGPASNTAPLOEQVFEETLEETIGYLYISAWPDSLP 420
|
QY 421 DLSVFONLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIHHTHLCFVHTV 480
|
DB 421 DLSVFONLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIHHTHLCFVHTV 480
|
QY 481 PWDQLFRNPQALLHTANRPEDECVEGLACHQLCARGHCWGPCTQCVNCSQFLRQEC 540
|
DB 481 PWDQLFRNPQALLHTANRPEDECVEGLACHQLCARGHCWGPCTQCVNCSQFLRQEC 540
|
QY 541 VEECRVLOGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHKDPFPCVARC 600
|
DB 541 VEECRVLOGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHKDPFPCVARC 600
|
QY 601 PSGVRPDLSYMIWKPDEEGACQPCINCTHSCVDLDDKGCPEORASPLTSIVSAVVG 660
|
DB 601 PSGVRPDLSYMIWKPDEEGACQPCINCTHSCVDLDDKGCPEORASPLTSIVSAVVG 660
|
QY 661 ILLVVVLGVVFGILIKRROOKIRKYTMRLLOQETELVEPLTPSGAMPNOQMRLKETEL 720
|
DB 661 ILLVVVLGVVFGILIKRROOKIRKYTMRLLOQETELVEPLTPSGAMPNOQMRLKETEL 720
|
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVGP 780
|
DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVGP 780
|
QY 781 YVSRLLGICLTSTVQLVLTQMPYGCCLLDHVRENRGRGLSGDQLLNWCQIAKGNLSYLEDVR 840
|
DB 781 YVSRLLGICLTSTVQLVLTQMPYGCCLLDHVRENRGRGLSGDQLLNWCQIAKGNLSYLEDVR 840
|
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIAKGNLSYLEDVR 900
|

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Qy 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDELGPASPLDSTFYRSILLEDGMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDELGPASPLDSTFYRSILLEDGMDGLVDA 1020
Qy 1021 EYLVPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLSPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLSPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPPPAFPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPPPAFPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 15

ID AAR39568 standard; Protein; 1433 AA.

AC AAR39568;

DT 07-FEB-1994 (first entry)

DE Sequence of c-erbB-2 tumour antigen.

KW Tumour antigen; c-erbB-2; glycoprotein.

OS Homo sapiens.

PN W09316185-A.

PD 19-AUG-1993.

PF 05-FEB-1993; 93WO-US01055.

PR 06-FEB-1992; 92US-0831967.

PA (CETU) CETUS ONCOLOGY CORP.
(CREA-) CREATIVE BIOMOLECULES INC.

PI Houston LL, Huston JS, Oppermann H, Ring DB;

XX WPI; 1993-272889/34.

DR N-PSDB; AAQ46083.

PT New single chain Fv polypeptide binding to C-erbB-2 tumour

PT antigen - for imaging or treating breast or ovarian cancer etc.

PS Disclosure; pages 48-54; 87pp; English.

XX c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.

XX SQ Sequence 1433 AA;

Query Match 98.2%; Score 6687; DB 14; Length 1433;

Best Local Similarity 98.3%; Pred. No. 0;

Matches 1234; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQTGCTDMKLRLPASPETHLDMLRHLVQGCVOVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQTGCTDMKLRLPASPETHLDMLRHLVQGCVOVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLAHNOVROVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLAHNOVROVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
Qy 121 DPLNNQYIKANSKFITGTELQRLSLTEILKGGVLIQRPOLCYQDTILMKDIFHKNQLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILMKDIFHKNQLA 180
Qy 181 LTLIDTVRSACHPCSPMKCSRCWGESSEDCOSLTRTVAGGCARCKGPLTDCCHQEC 240
Db 181 LTLIDTVRSACHPCSPMKCSRCWGESSEDCOSLTRTVAGGCARCKGPLTDCCHQEC 240
Qy 241 AAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCPLHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
Db 301 YNYLSTDVSGCTLVCPLHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEEITCYLYISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEEITCYLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRIIHLNGAYSILTQGLGISWLGRLSLRELGLGLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRIIHLNGAYSILTQGLGISWLGRLSLRELGLGLALIHNTHLFCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVEGLEACHOLCARGHCGWGPCTOCVNCQFSLRQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVEGLEACHOLCARGHCGWGPCTOCVNCQFSLRQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTFCGPEADOCVCAHYKDPFCVAVRC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTFCGPEADOCVCAHYKDPFCVAVRC 600
Qy 601 PSGVKPDLSPYMPKWPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660
Db 601 PSGVKPDLSPYMPKWPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660
Qy 661 ILLVWLVGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVWLVGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKYKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
Db 721 RKYKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLGSDLLNWCQIAKGSYLEDDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLGSDLLNWCQIAKGSYLEDDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Qy 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDELGPASPLDSTFYRSILLEDGMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDELGPASPLDSTFYRSILLEDGMDGLVDA 1020
Qy 1021 EYLVPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLSPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLSPSEEEAPRSLAPSEG 1080

| | | | | | |
|----|------|------------------------|--------------------------|--------------------------|------|
| Qy | 1081 | AGSDVFDGDLGMAAKGLOSLP | THDPSPLOQYSEDP | TVPLPSETDGYVAPLTCSPQPEYV | 1140 |
| Db | 1081 | AGSDVFDGDLGMAAKGLOSLP | THDPSPLOQYSEDP | TVPLPSETDGYVAPLTCSPQPEYV | 1140 |
| Qy | 1141 | NQPDVVRQPSPREGPLPAARPA | GATLERAKTILSPGKNGVVKDVFA | FGGAVENPEYLTPO | 1200 |
| Db | 1141 | NQPDVVRQPSPREGPLPAARPA | GATLERAKTILSPGKNGVVKDVFA | FGGAVENPEYLTPO | 1200 |
| Qy | 1201 | GGAAPQHPPPAFSPAFDNL | YYWDQPPPERGAPPSTFKGTPTA | ENPEYLGLDVVP | 1255 |
| Db | 1201 | GGAAPQHPPPAFSPAFDNL | YYWDQPPPERGAPPSTFKGTPTA | ENPEYLGLDVVP | 1255 |

Search completed: July 22, 2003, 08:40:28
Job time : 43.9774 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-149-163-12
Perfect score: 6810
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTFPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:.*
1: Piri:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 6734 | 98.9 | 1255 | 1 A24571 | protein-tyrosine k |
| 2 | 5935 | 87.2 | 1260 | 1 TVRTNU | protein-tyrosine k |
| 3 | 5931.5 | 87.1 | 1254 | 2 I48161 | p-185 precursor - |
| 4 | 3160 | 46.4 | 1210 | 1 GQHUE | epidermal growth f |
| 5 | 3135 | 46.0 | 1210 | 2 A53183 | epidermal growth f |
| 6 | 3109.5 | 45.7 | 1223 | 1 TVCHLV | epidermal growth f |
| 7 | 2975.5 | 43.7 | 1308 | 2 A47253 | epidermal growth f |
| 8 | 2684 | 39.4 | 1166 | 1 S06142 | protein-tyrosine k |
| 9 | 2423.5 | 35.6 | 1342 | 2 A36223 | kinase-related tra |
| 10 | 2339.5 | 34.4 | 1339 | 2 JCA387 | epidermal growth f |
| 11 | 1766.5 | 25.9 | 698 | 1 TVPVLV | protein-tyrosine k |
| 12 | 1703 | 25.0 | 604 | 1 TVYUOH | protein-tyrosine k |
| 13 | 1647 | 24.2 | 544 | 2 S35745 | protein-tyrosine k |
| 14 | 1645.5 | 24.2 | 1330 | 1 GQFFE | protein-tyrosine k |
| 15 | 1640 | 24.1 | 545 | 2 S00727 | kinase-related tra |
| 16 | 1623 | 23.8 | 540 | 2 B44776 | protein-tyrosine k |
| 17 | 1621 | 23.8 | 540 | 1 TVPVEB | protein-tyrosine k |
| 18 | 1530 | 22.5 | 644 | 2 A36325 | protein-tyrosine k |
| 19 | 1297 | 19.0 | 1323 | 2 E88257 | protein let-23 (im |
| 20 | 1297 | 19.0 | 1374 | 2 S70712 | protein-tyrosine k |
| 21 | 1208 | 17.7 | 1369 | 2 S70713 | epidermal growth f |
| 22 | 1162 | 17.1 | 1717 | 1 A45558 | epidermal growth f |
| 23 | 1141 | 16.8 | 527 | 2 A42032 | epidermal growth f |
| 24 | 990.5 | 14.5 | 843 | 2 A27131 | epidermal growth f |
| 25 | 806.5 | 11.8 | 346 | 2 S13807 | protein-tyrosine k |
| 26 | 754.5 | 11.1 | 311 | 2 S13808 | protein-tyrosine k |
| 27 | 725 | 10.6 | 1363 | 2 T43220 | insulin-like growt |
| 28 | 713 | 10.5 | 1382 | 1 INHUR | insulin receptor p |
| 29 | 705.5 | 10.4 | 1372 | 2 A34157 | insulin receptor p |

| | | | | | |
|----|-------|------|------|----------|--------------------|
| 30 | 705 | 10.4 | 1383 | 2 A36080 | insulin receptor p |
| 31 | 702 | 10.3 | 1607 | 2 T43212 | insulin-like growt |
| 32 | 698.5 | 10.3 | 1300 | 2 A36502 | insulin receptor-r |
| 33 | 684 | 10.0 | 1268 | 2 B36502 | insulin receptor-r |
| 34 | 669 | 9.8 | 1477 | 2 T18534 | protein-tyrosine k |
| 35 | 650 | 9.5 | 1367 | 1 IGHUR1 | insulin-like growt |
| 36 | 639 | 9.4 | 1371 | 2 A33837 | insulin-like growt |
| 37 | 634 | 9.3 | 1390 | 2 T30346 | insulin receptor - |
| 38 | 618.5 | 9.1 | 2148 | 1 A56081 | insulin receptor - |
| 39 | 615 | 9.0 | 2101 | 2 S57245 | insulin receptor (|
| 40 | 591.5 | 8.7 | 952 | 2 I50612 | protein-tyrosine k |
| 41 | 591 | 8.7 | 987 | 2 A54092 | protein-tyrosine k |
| 42 | 589.5 | 8.7 | 977 | 2 S49004 | tyrosine kinase Mp |
| 43 | 588 | 8.6 | 1114 | 1 S05582 | protein-tyrosine k |
| 44 | 586 | 8.6 | 1091 | 2 S33596 | protein-tyrosine k |
| 45 | 584.5 | 8.6 | 976 | 2 A36355 | protein-tyrosine k |

ALIGNMENTS

RESULT 1

A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N:Alternate names: c-erbB-2 protein precursor; kinase-related transforming protein e
C:Species: Homo sapiens (man)
C:Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.
Nature 319, 230-234, 1986
A:Title: Similarity of protein encoded by the human c-erbB-2 gene to epidermal growth
A:Reference number: A24571; MUID:86118663; PMID:3003577
A:Accession: A24571
A:Molecule type: mRNA
A:Residues: 1-1255 <YAM>
A:Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198
R:Semba, K.; Kanata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid
A:Reference number: A25491; MUID:86016729; PMID:2995967
A:Accession: A25491
A:Molecule type: DNA
A:Residues: 737-1031 <SEM>
A:Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282
R:Cousens, L.; Yang-Peng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg
Science 230, 1132-1139, 1985
A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro
A:Reference number: A44188; MUID:86070181; PMID:2999374
A:Accession: A44188
A:Molecule type: DNA
A:Residues: 740-910 <COU1>
A:Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989
A:Accession: B44188
A:Molecule type: mRNA
A:Residues: 1-517, 'PALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A:Cross-references: GB:M11730; NID:G183986
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A:Reference number: I59509; MUID:85272597; PMID:2992089
A:Accession: I59509
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 832-909 <REX>
A:Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio
A:Reference number: I57622; MUID:87286898; PMID:3039351
A:Accession: I57622
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAAS8637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <BE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <BE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.9%; Score 6734; DB 1; Length 1255;
Best Local Similarity 98.8%; Pred. No. 7.4e-268;
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASQTGCTDMKLRLPASPETHLDMLRHLHLYQGCVVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASQTGCTDMKLRLPASPETHLDMLRHLHLYQGCVVQGNL 60
QY 61 ELYLPTNASLSFLQDIQEVQGVYLIHANNVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
DB 61 ELYLPTNASLSFLQDIQEVQGVYLIHANNVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDQYIKANSKF 180
DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDQYIKANSKF 180
QY 181 IGTENLRSRACHPCSPMKGSRGWGESSEDCOSLTRTCVAGGCARCKGLPTDDCHEQC 240
DB 181 LTLLDNRSPACHPCSPMKGSRGWGESSEDCOSLTRTCVAGGCARCKGLPTDDCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNHOEVTAEQGTQRCCKSPCARVCYGLGWEHLREVRVTSAN 360
DB 301 YNYLSTDVGSCTLVCPLNHOEVTAEQGTQRCCKSPCARVCYGLGWEHLREVRVTSAN 360
QY 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVPETLEEITGYLIYISAWPDSLP 420
DB 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVPETLEEITGYLIYISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRIHLHNGAYSITLQGLISWGLSLRSLRGLSGLALIHNTLHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRIHLHNGAYSITLQGLISWGLSLRSLRGLSGLALIHNTLHLCFVHTV 480
QY 481 PWDLFRNPHOALLHTANRPEDCEVGEGLACHOLCARGHCWGPQTQVCNCSOFLRGQEC 540
DB 481 PWDLFRNPHOALLHTANRPEDCEVGEGLACHOLCARGHCWGPQTQVCNCSOFLRGQEC 540
QY 541 VEECRVLQGLPREYVYVNAHCLPCHPECQPGQNGSVTCFGEADOCVACAHYKDPFCVCARC 600
DB 541 VEECRVLQGLPREYVYVNAHCLPCHPECQPGQNGSVTCFGEADOCVACAHYKDPFCVCARC 600
QY 601 PSQVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTISIISAVVG 660
DB 601 PSQVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTISIISAVVG 660

DB 601 PSQVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTISIISAVVG 660
QY 661 ILLVVLGVVFGILLIKRQOKIRKYIMRLLQBTETELVEPLTPSGAMPNQAMRILKTEL 720
DB 661 ILLVVLGVVFGILLIKRQOKIRKYIMRLLQBTETELVEPLTPSGAMPNQAMRILKTEL 720
QY 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
DB 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVLTQMLPYGCLLDHVRENRGLSGODLLNMCQIAKGSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVLTQMLPYGCLLDHVRENRGLSGODLLNMCQIAKGSYLEDVR 840
QY 841 LVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRFT 900
DB 841 LVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMVKCWM 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMVKCWM 960
QY 961 IDSECRPRRELVSFESRMARDQRFVITQNEDLGPASPLDSTFYRSLLDEDDMGDLVDA 1020
DB 961 IDSECRPRRELVSFESRMARDQRFVITQNEDLGPASPLDSTFYRSLLDEDDMGDLVDA 1020
QY 1021 EYLVVPOQGFCCPDPAAGAGVHHRSSSTSGGGDLTLGLEPSEEEAPRSLAPSEG 1080
DB 1021 EYLVVPOQGFCCPDPAAGAGVHHRSSSTSGGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYV 1140
QY 1141 NQPDVVRPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVAFPGGAVENPEYLTPO 1200
DB 1141 NQPDVVRPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVAFPGGAVENPEYLTPO 1200
QY 1201 GGAAPOHPHPPAFSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
DB 1201 GGAAPOHPHPPAFSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohe
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals r
2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663, 'V', 665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>


```
F:558-680/Domain: transmembrane #status predicted <TM>
F:732-998/Domain: protein kinase homology <KN>
F:731-739/Region: protein kinase ATP-binding motif
F:711_191_263_535_576_634/Binding site: carbohydrate (AS)
F:691/Binding site: phosphate (Thr) (covalent) #status
F:788/Active site: Lys #status predicted
F:882_1227_1253/Binding site: phosphate (Tyr) (covalent)
```

| Query Match | 87.2% | Score 5935 | DB 1 | Length 1260 |
|-----------------------|--------------|------------------|----------------|-------------|
| Best Local Similarity | 87.1% | Pred. NO. 3e-235 | | |
| Matches 1095 | Conservative | 51 | Mismatches 109 | Indels 2 |
| Gaps 2 | | | | |

| | | | | | | | | | |
|----|-----|--------------------------|-----------------------------|--------------|--------------|-----------------------|------------------------------|-----------------|-----|
| Qy | 1 | MELAAALCRWGLJLLALLPPGAAS | QVCTGTDMKRLR | PAS | PETHDMLRHL | YQCVQVQGNL | 60 | | |
| Db | 4 | MELAAWCRWGFLLALLPPGIA | QVCTGTDMKRLR | PAS | PETHDMLRHL | YQCVQVQGNL | 63 | | |
| Qy | 61 | ELTYLPTNASL | FLQDIOEVGYVLI | AHNOVQVPLQRL | RIVRG | TQSLFEDNWALAVLNG | 120 | | |
| Db | 64 | ELTYVVPANAS | LUSFLQDIOEVGYMLIAHNOVKR | VPQLRLR | IVRG | TQSLFEDKFKYALAVLNR | 123 | | |
| Qy | 121 | DPLANNTPVT | -GASPGGLURELQLRSLTEILKGGVLI | ORNPOL | CYQDTIL | WKDOYIKANSK | 179 | | |
| Db | 124 | DPQDNVAAS | TGRTPEGLRELQLRSLTEILKGGVLR | IGNPQL | CYQDMVL | WKDVPFRNQOL | 183 | | |
| Qy | 180 | FIGITELNRS | RACHPCSPMKGSR | CWGSSEDCQSL | TRTV | CAGGCARCKGPLPTDCCHEQ | 239 | | |
| Db | 184 | APVDIDTNR | SACPPCAPACKDNHCWGS | PEDCQIL | TG | TICTSGCARCKGR | LPTDCCHEQ | 243 | |
| Qy | 240 | CAAGCTGPKHSDCLAC | LHPNHSIGICEL | HPALV | TYNTDT | FESMPNPEGRYTFGASCVTAC | 299 | | |
| Db | 244 | CAAGCTGPKHSDCLAC | LHPNHSIGICEL | HPALV | TYNTDT | FESMHNPEGRYTFGASCVTTC | 303 | | |
| Qy | 300 | PYNLYSTDVGS | CTLVCP | LHNQEV | TAEDGTQ | RCEKSKPCAR | CYGLGMEHLREVAVTS | 359 | |
| Db | 304 | PYNLYSTEVGS | CTLVCP | PNNQEV | TAEDGTQ | RCEKSKPCAR | CYGLGMEHLRGARITSD | 363 | |
| Qy | 360 | NIQFAGCKK | ITFGLAF | PESFDGDPAS | NTAP | LOPELQV | FETLBEITCYLYISAWPDSL | 419 | |
| Db | 364 | NVQFDFGCKK | ITFGLAF | PESFDGDPSS | GIAPLR | PELQV | FETLBEITCYLYISAWPDSL | 423 | |
| Qy | 420 | PDLVSFQNLQV | IRGRI | LHNGAYS | LTLOG | LIGISW | LGRLSIRBELSGGLAIHHNTHLCFVHT | 479 | |
| Db | 424 | RDLVSFQNLRI | IRGRI | LHDGAYS | LTLOG | LIGIHS | LGRLSIRBELSGGLAIHRRNAHLCFVHT | 483 | |
| Qy | 480 | VPDQLFRNPHQALLHT | ANRPEDE | -CVGEG | LACHOLCARGHC | WGP | GPTQCVNCSFLRGO | 538 | |
| Db | 484 | VPDQLFRNPHQALLHSGNR | PEEDLCVSSGL | VCNCSLCAHGC | WGP | GPTQCVNCSFLRGO | 543 | | |
| Qy | 539 | ECVBEERVQLPREY | VNARHCLPCH | PECQ | PONGSVTCF | GPEADOC | VCACAHYKDPFFCVA | 598 | |
| Db | 544 | ECVEERVWKGLPREY | VSDKRCLPCH | PECQ | PONSSETCF | GSEADCAACAHYKDNSSCVA | 603 | | |
| Qy | 599 | RCPSGVKPDLSYMP | IWKFPDDEGAC | QPCP | INCTHSCVDL | DKGCP | PAEORASPLTSIVSAV | 658 | |
| Db | 604 | RCPSGVKPDLSYMP | IWKYPDEBEG | QCPCP | INCTHSCVDL | DERGCP | PAEORASPVTEIATV | 663 | |
| Qy | 659 | VGILLVVVLGVV | GILIKRRQ | KIRKY | TMRLLO | ETLVEPLTP | SGAMPNQAOMRILKET | 718 | |
| Db | 664 | EGVLLFLILVVV | GILIKRRQ | KIRKY | TMRLLO | ETLVEPLTP | SGAMPNQAOMRILKET | 723 | |
| Qy | 719 | ELRKVKVLGSGA | FGFTVYKGIWI | PDGENVKI | PVAIKV | LRENTS | SPKANKEILDEAYVMAGV | 778 | |
| Db | 724 | ELRKVKVLGSGA | FGFTVYKGIWI | PDGENVKI | PVAIKV | LRENTS | SPKANKEILDEAYVMAGV | 783 | |
| Qy | 779 | SPYVSRLIGICLT | STVQLVTQ | LMPYGCLLD | HVREN | RGRGLSGQD | LLNWCQI | AKGMSYLE | 838 |
| Db | 784 | SPYVSRLIGICLT | STVQLVTQ | LMPYGCLLD | HVREH | RGRGLSGQD | LLNWCQI | AKGMSYLE | 843 |
| Qy | 839 | VRLVHRDLAARN | NVLKSPNHVKI | ITDFGL | ARLLD | IDETEH | HADGGKVP | IKKMALESILRRR | 898 |
| Db | 844 | VRLVHRDLAARN | NVLKSPNHVKI | ITDFGL | ARLLD | IDETEH | HADGGKVP | IKKMALESILRRR | 903 |
| Qy | 899 | FTHOSDVMSYGV | TWELMTFGAKPYD | GI | IPAREI | PDLEK | GRLLPOP | PICTIDTVIMIMVKC | 958 |

| | | | | | | | | | | | | |
|------|------------------------------|----------|--------------------------|-----------------------|-----------|--------|------|--------------|--------------|---------|-------|------|
| 904 | FTHQSDVMSYGVTVWMLTFGAKPYDGI | PAREI | PDLEKGERLPQPICTIDVYIMVKC | 963 | | | | | | | | |
| 959 | WMIDSECRPFRELVSFESRMARDPORFV | YIQNE | DLGPA | SLDSTFYRSLLEDDDDMGDLV | 1018 | | | | | | | |
| 964 | WMIDSECRPFRELVSFESRMARDPORFV | IQNE | DLGPS | PMDSTFYRSLLEDDDDMGDLV | 1023 | | | | | | | |
| 1019 | DAZEYLVPOQGFCDPAPGAGGMVHRHS | STRSG | GGDLTLGL | SESEEA | PRSLAPS | 1078 | | | | | | |
| 1024 | DAZEYLVPOQGFSPDPTCTGSTAHRHRS | STRSG | GGELTLGL | SESEEG | PRSLAPS | 1083 | | | | | | |
| 1079 | EGAGSVDFDGLMGAAKGLSLPDP | QLQRYSED | TVLP | SETDGVYAP | LTC | SPOE | 1138 | | | | | |
| 1084 | EGAGSVDFDGLAMGVTKGLQSLPS | HDLSPL | QLQRYSED | TVLP | PETDGVYAP | LAC | SPOE | 1143 | | | | |
| 1139 | YVNPQDVRPOPSPRSGPLPAAR | PAGAT | LERA | KTLS | PGKNGV | VVKD | YAF | CGGAVENPEYLT | 1198 | | | |
| 1144 | YVNPQSEVQOPPLTPGEP | LPVP | PAGAT | LER | KTLS | PGKNGV | VVKD | YAF | CGGAVENPEYLV | 1203 | | |
| 1199 | POGGAAPQPHPPAF | SPAF | DNLYW | YQDD | PPER | GAPP | STFK | GTTA | BNPEYLG | LDVVP | 1255 | |
| 1204 | PRGCTASPHPSAF | SPAF | DNLYW | YQNS | SE | QGP | PPSP | NEG | TPTA | BNPEYLG | LDVVP | 1260 |

RESULT 3

I48161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.;
Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:G493236; PIDN:BAAO3801.1; PID:G747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: AIP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

| | | | | |
|-----------------------|--------------|--------------------|----------------|-------------|
| Query Match | 87.1% | Score 5931.5 | DB 2 | Length 1254 |
| Best Local Similarity | 86.9% | Pred. No. 4.2e-235 | | |
| Matches 1091 | Conservative | 59 | Mismatches 104 | Indels 1 |
| | | | | Gaps 1 |

[illegible]

```
QY 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVFETLEBITGVLYLISAMPDLSLP 420
DB 361 IQEFAGCKKIFGSLAFLPESFDGDPNFGSSGIAPLTPPEQLQVFETLEBITGVLYLISAMPDLSH 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGSLWLSRLSRLGSLALHINHLFCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGSLWLSRLSRLGSLALHINHLFCFVHTV 480
QY 481 PWDQLFRPHQALLHTANRPEDECVEGLACHQLCARGHCWGPPTQCVCNCSQFIRGQEC 540
DB 481 PWDQLFRPHQALLHTANRPEDECVEGLACHQLCARGHCWGPPTQCVCNCSQFIRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
DB 541 VKECRVWKLPREYVNGKILCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
QY 601 PSGVKPDLISYMPIWKPDEEGACQPCINCHTSCVDLDDKGPABORASPLTSIVSAVVG 660
DB 601 PSGVKPDLISYMPIWKPDEEGACQPCINCHTSCVDLDDKGPABORASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
DB 661 ILLVLVIGVVGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
DB 721 RKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGLICTSTVOLTMPLPYGCLLDHVRENRLGSGQDLNNCMQIAKMSYLEDV 840
DB 781 YVSRLLGLICTSTVOLTMPLPYGCLLDHVRENRLGSGQDLNNCMQIAKMSYLEDV 840
QY 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHYHAGCKVKIPKMALESILRRFT 900
DB 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHYHAGCKVKIPKMALESILRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCM 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCM 960
QY 961 IDSECRPRFRELVSFSEARMARDQRFVVIQNSDLGSPDLSTFYRSLLLEDMDGLVDA 1020
DB 961 IDSECRPRFRELVSFSEARMARDQRFVVIQNSDLGSPDLSTFYRSLLLEDMDGLVDA 1020
QY 1021 BEYLVPOQGFCDPAPGAGVHHRSSSTRSGGDLTLGLEPSESEAPSLAPSEG 1080
DB 1021 BEYLVPOQGFCDPAPGAGVHHRSSSTRSGGDLTLGLEPSESEAPSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLTPHDPSPQLQRYSEDTVPPLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFEGLGNGATKQPSISPRDLSPQLQRYSEDTVPPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NOPDVRRPQPSRPGPLPAARAGATLERAKTSLSPKQGVVWDVAFAGAVENPEYLP 1200
DB 1141 NOPEVRPQPLPTEGRLPPVRPAGATLERKPTLSFGKGVVWDVFTFGAVENPEYLP 1200
QY 1201 GGAAPQHPPPAFSAFDNLYYWDQDPPRGAPPSTFKGTPTAENPEYVGLDVPV 1255
DB 1201 GGSASQPH-PPALCAFDNLYYWDQDPPSERGSPNFTFGTPTAENPEYVGLDVPV 1254
```

RESULT 4

GCHUE

epidermal growth factor receptor precursor - human

N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Homo sapiens (man)

C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999

C:Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143; A33

R:Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y

rg, P.H.

Nature 309, 418-425, 1984

A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <ULL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g5757924
A:Note: The authors translated the codon AAG for residue 540 as Aen
R: Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M1234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R:Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ulrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termi
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A:Experimental source: carcinoma cell line A431-7
R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.;
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RN
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 'RCANRA', 150-187, 'KSVIQAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-
'798-799, 'TD', 802-811, 'R', 813-942 <XUY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF re
R:Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma,
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplificati
A:Reference number: A43615; MUID:84196372; PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A:Title: Identification of residues in the nucleotide binding site of the epidermal g
A:Reference number: A60143; MUID:85182650; PMID:2985580
A:Accession: A60143
A:Molecule type: protein
A:Residues: 740-744, 'X', 746-747 <RUS>
R:Mróczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and supe
A:Reference number: A38023; MUID:84191554; PMID:6325948
A:Contents: annotation; receptor activity

A>Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R.Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
Cell 59, 33-43, 1989
A:Title: Functional independence of the epidermal growth factor receptor from a domain
A:Reference number: A3331; MUID:90003233; PMID:2790960
A:Contents: annotation: internalization signal
C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor
C:Genetics:
A:Gene: GDB:EGFR
A:Cross-references: GDB:120610; OMIM:131550
A:Map position: 7p12.3-7p12.1
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-1210/Product: EGF receptor #status predicted <MAT>
F:25-645/Domain: extracellular #status predicted <EXT>
F:75-300/Domain: EGF receptor extracellular domain repeat <EE1>
F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>
F:669-1210/Domain: transmembrane #status predicted <TM>
F:646-668/Domain: EGF receptor extracellular domain repeat <EE2>
F:669-1210/Domain: intracellular #status predicted <INT>
F:710-975/Domain: protein kinase homology <KIN>
F:718-726/Region: protein kinase ATP-binding motif
F:999-1046/Region: coated-pit mediated internalization signal
F:1047-1210/Region: inhibitory
F:128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic
F:745/Active site: Lys #status experimental

Query Match 46.4%; Score 3160; DB 1; Length 1210;
Best Local Similarity 49.8%; Pred. No. 5,4e-122;
Matches 631; Conservative 177; Mismatches 348; Indels 112; Gaps 23;

| | | | |
|----|-----|--|-----|
| QY | 11 | LLALLPGAA--STQVCTGTDMKRLPASPETHLDMLRLHYOGCOVGNLELTPLTN | 68 |
| DB | 14 | LLAALCPASRALEKKVCQGTNKLTQLGTGFHFLSLQRMFNCEVILGNLEITYVQRN | 73 |
| QY | 69 | ASLSFLDIOEQVGYVLIHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNGPLNNTTP | 128 |
| DB | 74 | YDLSEFKTIOEVAGYVLIHNTVERIPLNQLIRGNMYIENSYALAVLSNYD----- | 126 |
| QY | 129 | VTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYDITLWKDQIKANSKFIGITEL-- | 186 |
| DB | 127 | ---ANKTGLKELPMRLQLBILHGAVRFSNNPALCNVESIQWRD---IVSSDFLSNMDP | 180 |
| QY | 187 | -NRSRACHPCSPMKGRSCWGSSEDCQSLRTVTCAGCA-RCKGPLPTDCHEQCAAGC | 244 |
| DB | 181 | QNHLSGCRQCDPSCPNCSGWCAGEENCQKLTKIIQAQCSGRGKSPSDCHNQCAAGC | 240 |
| QY | 245 | TGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTEGASCVTACPNYL | 304 |
| DB | 241 | TGPRESDCLVCRKFRDEATCKDTCPLMLYNTTYQMDVNPPEGKVSFGATCVKCPRNVY | 300 |
| QY | 305 | STDVGSCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSANIQEF | 364 |
| DB | 301 | VTDHGSCVRACADSYEM-EDGVKCKCEGCRKVCNGIGIGEFKDSLSINATNIKF | 359 |
| QY | 365 | ACKKIFGSLAPLPSFDGDPASNTAPLOEQVFTLEETIGYLIYSAMPDLSV | 424 |
| DB | 360 | KNCTSIGSLHLTPVAFRGDSFTHTPPLDQBLDKTKVKEITGFLIQAQWPNRTDLHA | 419 |
| QY | 425 | FONLQVIRGRILHNGAYSITLQGLTISGLRSLRELGLALIHNNHLCFVHTVPDQ | 484 |
| DB | 420 | FENLEIIRGTQHQOFSLAVALNLSITSLGLSLKEISDGVLIISGNKRLCVANTINMK | 479 |
| QY | 485 | LFRNPQALLHTANRPEDEVCEGLACHOLCARGHCWGFQTCVNCOSFRLRGQCEVBC | 544 |
| DB | 480 | LFGTSGQTKIISNRGENSKATQVCHALCSPEGCWGPEPRDCVSCRNVSRGRCVDC | 539 |
| QY | 545 | RVLOQLPREYNARHCLPCHPCQPNQSVTCFGEADOCVACAHYKDPFFCVARCPGV | 604 |
| DB | 540 | KLLEGEPEFVENSECIOCHPECLFOAMNITCTGRPDNCTIOCAHYIDGPHCVKTC | 599 |
| QY | 605 | KPDLSPMPKPFDEEGACQPCPINCTHSCVDLDKGCAPAEQASPLTSIVSAVVG--I | 661 |

| | | | |
|----|------|--|------|
| Db | 600 | MEENNTL-VWKYADAGHVCHLCHPNCTYCTGTGLEGCTNGPKIP--SIATGMVGALLL | 656 |
| QY | 662 | LLVWLGVVFGILIKRQOKIRKYTMRRLLOETELVEPLTPSGAMPNQAMRILKETELR | 721 |
| Db | 657 | LLVVALGIG--LFMRRRHIVKRTLRLLORELVEPLTPSGEAPNQALLRLKETEFK | 713 |
| QY | 722 | KVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSPY | 781 |
| Db | 714 | KIKVLGSAFGTVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPH | 773 |
| QY | 782 | VSRLIGICLTSTVOLVTQMLMPGCLLDHVRNRLGSLDNLNWCQIAKMSYLEVDRL | 841 |
| Db | 774 | VCRLIGICLTSTVOLITQMLMPGCLLDYVREHKDNIGSQYLLNWCQIAKMSYLEDRRL | 833 |
| QY | 842 | VHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWMALLESILRRFTH | 901 |
| Db | 834 | VHRDLAARNVLKTPQHVKITDFGLAKLGAEEKEYHAEGGKVPKWMALLESILHRIYTH | 893 |
| QY | 902 | QSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPICTIDVYIMVAKCMI | 961 |
| Db | 894 | QSDVMSYGVTVWELMTFGSKPYDGIPASISSILEKGERLPQPICTIDVYIMVAKCMI | 953 |
| QY | 962 | DSECPREPRELVSERSMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDMDGDLVDA | 1020 |
| Db | 954 | DADSRPKFRELIEFSEKWARDQRYLVIQGDERMHLPSPDTSNFYRALMDEEDMDVDVA | 1013 |
| QY | 1021 | BEYLVPOQGFCCPDPAAGAGGVHHRHSSTRSGGDLTLGLERPEEAPRSLAPSEG | 1080 |
| Db | 1014 | DYLIPOQGF-----SSPSTRPLLSLS | 1039 |
| QY | 1081 | AGSDVFDGLGMAAKGQSLPDPSPLOQRYSEDPVLPSET--DGYPVAPLTCSPQE | 1138 |
| Db | 1040 | ATSN--NSTVACIDRNLGQCPKEDSFLOQRYSSDPTGALTEDSIDTFL-----PVPE | 1091 |
| QY | 1139 | YUNQDVPQPPSPREGPLPAARPAGATLERAKTLSPGKGVGVKDVAFEGGAVENPEYL- | 1197 |
| Db | 1092 | YINQ-SVPKRPAQSVQNPVYHNOPLN-----APSRDPYQD--PHSTAVGNPEYL | 1140 |
| QY | 1198 | TPQGAAPQPHPPAFSPAFDNLXYWQ-----DP-----PERGAPSPFKGTP | 1241 |
| Db | 1141 | TVQ-----PTCVNSTFDSPAHWAQSGHQISLDNPDYQODFFPKAKPNIGI | 1190 |
| QY | 1242 | TAENPEYL 1249 | |
| Db | 1191 | TAENAYEL 1198 | |

RESULT 5
A53183
epidermal growth factor receptor precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
C:Accession: A5318; A43818; S24942; A28941; S45325; I49643
R:Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Jenkins, N.A
Genes Dev. 8, 399-413, 1994
A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
A:Reference number: A53183; MUID:94170986; PMID:8125255
A:Accession: A53183
A:Molecule type: mRNA
A:Residues: 1-1210 <LUE>
A:Cross-references: GB:U03425
R:Avi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding
A:Reference number: A43818; MUID:91232866; PMID:2030916
A:Accession: A43818
A:Molecule type: mRNA
A:Residues: 1-714 <AVI>
A:Cross-references: GB:X59698
R:Eisinger, D.P.; Serrero, G.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24942
A:Accession: S24942

A:Molecule type: mRNA
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
 A:Cross-references: EMBL:Z12608
 R:Reisermann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941; MUID:88330814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000; 1002-1009,
 R:hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: S45325
 A:Accession: S45325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971, 'K', 973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
 A:Reference number: I49643; MUID:91126380; PMID:7678348
 A:Accession: I49643
 A:Status: translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 12-20, 22-132 <RES>
 A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
 C:Genetics:
 A:Gene: EGFR
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-670/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:630,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:7:0-728/Region: protein kinase ATP-binding motif
 F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 46.0%; Score 3135; DB 2; Length 1210;
 Best Local Similarity 49.6%; Pred. No. 5.7e-121;
 Matches 629; Conservative 171; Mismatches 366; Indels 102; Gaps 21;

Qy 11 LLLALLPPGAA--STQVCTGTDKMLRLPASPETHLMLRLHYQCCVQVQGNLELYLPTN 68
 Db 14 LLTALCAAGALEEKVKCGTNSRLTQGTGTFEDFLSLQRMYNCEVYVGNLEITYVQRN 73
 Qy 69 ASLSFLQDIOEVQYVLIHANVROVPLQRLRIVRGTPOLFEDNVALAVLNGDPLNNTTP 128
 Db 74 YDLSFLKTIQEVAGYVLIANTVERIPLENLQIRGNALYENTYALILSN----- 124
 Qy 129 VTGASPGGLRELQRLSITEILKGGVLIQRNPOLCYQDTILWKDQYIKANSKFIGITELNR 188
 Db 125 -YGNTRTGLRELPMRLQELIGAVRFESNPNILCNMDTIQWRDIVQNVFNSNMDLQSH 183
 Qy 189 SRACHPCSPCKSRCHWGESSEDCOSLTRVVCAGGCA-RCKGPLPTDCHEQCAAGCTGP 247
 Db 184 PSSCPKCDPSCPNCGSCWGGGEENCKLTKIICAAQCSHRRCGRSPDSCCHNQCAAGCTGP 243
 Qy 248 KHSDDLACLHFNHSGICELHCPALVTYNTDTFESMNPREGYTFGASCVTACPVNYLSTD 307
 Db 244 RESDCLVCQKQFQEAETCKDTPPLMLYNPTTYQNDVNPGEKYSFGATCVKKCPRNVYVD 303
 Qy 308 VGSCTLVCPHUNQVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVAVTSANIQEFAGC 367
 Db 304 HGSCVACAGPDYVEV-EEDGIRKCKKCDGCRKVCNGIGIGEFKDTLSINATNISKFKYC 362
 Qy 368 KKIYFGLAPFESFDGPASNTAPLQPEQLQVETLEETTCYLYISAWPDLSDLVQFN 427
 Db 363 TAISGDHLPLVAFKGDGFSFRTPPLDPRELEILKTKVEITGTFLLIQAWPDNWTDLHAFEN 422

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C;Species: Gallus gallus (chicken)

C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C;Accession: A27720; A00643

R;Tax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Ventr

Mol. Cell. Biol. 8, 1970-1978, 1988

A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in

A;Reference number: A27720; MUID:88261272; PMID:3260329

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R.Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1995
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A06643; MUID:85228222; PMID:2988784
A:Accession: A06643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <EXT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136-202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.7%; Score 3109.5; DB 1; Length 1223;
Best Local Similarity 48.6%; Pred. No. 6.3e-120;
Matches 630; Conservative 175; Mismatches 347; Indels 145; Gaps 25;

QY 8 RMGLLALLPPGAA-----STQVCTGTDMKRLPASPETHLDMLRLHYOGQVQGNLE 61
DB 13 RGAALVLLLLGVALCSAVEEKVCGCTNNKLTQLGHVEDHTSLORMNNECVLSNLE 72
QY 62 LTYLPTNASLSFLQDIQEOGVYLIYAHNOVRQVQLRIRVRGTOLFDENYALAVLDNGD 121
DB 73 ITYVEHNRDLTELKTIQEVAGYVLIYALNMVDVPLENLQIRGNVLYDNSFALVLSNVH 132
QY 122 PLNNITPVGASGGLRELQSLRSLTEILKGGVLIQBNPOLCYQDTILWKDQYIKANSKFI 181
DB 133 -MNKTO-----GLRELPLKRLSELNGGVKISNNPKLCNMNDTVLMND-IIDTSRKPL 182
QY 182 GITEL-NRSRACHPCSPMCKSGRSGESSEDCQSLTRTVACGGCA-RCKGPLPTDCCHQ 239
DB 183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQCSGRCKRVPSCCHNQ 242
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 243 CAAGCTGPRSDCLACRFRDDATCKDTCPLVLYNPTTYQMDVNPPEGKYSFGATCVR 302
QY 300 PYNLYLTDVGSCTLVCLPLHNOBVTAEQTCREKSKPCARVCYGLGMEHLREVRATSA 359
DB 303 PHNYVYTDHGSCVRSNTDTYEV-BENGVRKCKCKDGLCSKVCNGIGIGELGKILSINAT 361
QY 360 NIOEFAGCKKIFGSLAFSPESDGPASNTAPLOEQLOVFTLEETITGLYIYISAWPDSL 419
DB 362 NIDSFNCKTKINGDVSILPVALGDAFTKLPDKLDVFTVKEISGFLLIQAWPDNA 421
QY 420 PDLVSFQNLQVIRGRILHNGAYSLTQGLIGISWGLRLSLRELGSGLAIHHTHLCFVHT 479
DB 422 TDLVAFENLEIIRGTRTKHQGYSLAVNLKIQSLGLRSLKEISDGLIIMKNKILCYADT 481
QY 480 VPMDOLERNPHQALLHTANPEDECVGECGLACHQICARGHCWGPPTQVCNCSQFLRQE 539
DB 482 MNWRSLFATQSQTKI IQNRNKNCDTADRHVCDPLCSVDGCMGPGPFHCFSCRFFSRQKE 541
QY 540 CVEECRVLQGLPREYVYNARHCLPCHPECPQNG---SVTCFGPEADQCACAHYKDPFC 596

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A:Title: Ligand-specific activation of HER4/pl80erbB4, a fourth member of the epiderm

A:Reference number: A47253; MUID:93189574; PMID:8383326

A:Accession: A47253

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1308 <PLO>

A:Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360

A>Note: sequence extracted from NCBI backbone (NCBIP:126842)

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; growth factor receptor

F:716-981/Domain: protein kinase homology <KIN>

F:724-732/Region: protein kinase ATP-binding motif

| | | | | | |
|-----------------------|------|--|-------------------|-------------|--------------|
| Query Match | | 43.7% | Score 2975.5; | DB 2; | Length 1308; |
| Best Local Similarity | | 45.2% | Pred. No. 2e-114; | | |
| Matches 609; | | Conservative 185; | Mismatches 379; | Indels 173; | Gaps 28; |
| QY | 9 | WGULLALLPGAA-----STOVCTGTDKMLRPLASPETHLDMLRHLVQGCQVVOGNEULTY | 64 | | |
| Db | 8 | WVWSLVAAAGTVQPSDSQVCAGTENKLSLSLEQOYRALRKYYENCEVVMGNLEITS | 67 | | |
| QY | 65 | LPTNASLFLQDIQEVGYVLIAHNOVROVPLQRLRIVRGTOQLFEDNYALAVLDNGDPLN | 124 | | |
| Db | 68 | IEHRDLSFLRSREVVTGYVLVALNQRYLPLENLRIIRGTLYKLYEDRYALAIFLNVRKDG | 127 | | |
| QY | 125 | NTTPVTGASPGGLREQLRSLTEILKGVLIQIRNPQLCYQDTILWKDQYIKANSKFIGIT | 184 | | |
| Db | 128 | NF-----GLQELGLKNLTELNGVGVVDQNKFLCYADTIHQWDIVRNPWPSNLTIV | 178 | | |
| QY | 185 | ELNRSRACHPCSPMKGSRGWCAGSSEDCQSLTRTVCAAGC-ARCKGLPLPTDCHEQCAAG | 243 | | |
| Db | 179 | STNGSGCGRCHKSCTG-RCMGPTENHCOTLTRTVCAEQCDGRCYGYVSDCCHECAGG | 237 | | |
| QY | 244 | CTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFEGASCVTACPNY | 303 | | |
| Db | 238 | CSGPKOTDCACMNFDSGACVTCQPOTFVFNPTTFOLEHNFNAKYTYGAFCKKCPHNF | 297 | | |
| QY | 304 | LSTDVGSCTLVCPHLHNOEVTAEADGTQRCCEKSKPCARVCYGLGMEHLREVRVTSANIOE | 363 | | |
| Db | 298 | V-VDSSSCVCRACPSKMEV-EENGIKWCKPCTDICPKACDGIQTGSLMSAQTVDSNNIDK | 355 | | |
| QY | 364 | FACKKIFGLAFELPESFGDGPASNTAPLOEQLOVFPETLEBITYIYISAWPDSLPLDS | 423 | | |
| Db | 356 | FINCTKINGNLIFLVGTIHDGPDYNAIEADPEKLNVFRTVREITFGLNIQSWPNNMTDFS | 415 | | |
| QY | 424 | VFNQLQVIRGILRHNGAYSITLQGLISWLGSLRSLGSLALIHNTHLCFVHTVPWD | 483 | | |
| Db | 416 | VFNLTVIGRVLVYGLSLILKQOGLTSLOFGLKEISAGNIYITDNSLVCYHTINWT | 475 | | |
| QY | 484 | QLFRNPQALLHTANRPEDECVGEGLACHOLCARGHCWGPPTQCVCNCSQFLRQECVEE | 543 | | |
| Db | 476 | TLFSTINQIRVIRDNRAENKAETAGMVCNHLCSDDGCGWGPQDCLSCRRFSRGRICIES | 535 | | |
| QY | 544 | CRVLOGLPRYVYNARHCLPCHPECOP-QNGSVTCFGEADQCACAHYKDPFPCVACRPS | 602 | | |
| Db | 536 | CNLYDGEFREFFENGISICEVDCPOCEKMEDEGLLTCHGPGPNCTKCSHFKGPNCKVEKCPD | 595 | | |
| QY | 603 | GVPKDLISYMIWIPFPDEGACQPCPINCTHSCVDLDDKGC-----PAEQRASPL | 651 | | |
| Db | 596 | GLOGANSF--IFKYADPDRECHCPNCTGCGNPTSHDCIYYPWTHGSHLPOHAR-TPL | 652 | | |
| QY | 652 | TSIVSAVV-GILLVVLGVVFGILIKRROOKIRKYTWRRLLQETELVEPLTPSGAMPNOA | 710 | | |
| Db | 653 | --IAAGVIGGLFVLVIVGLTFVAVVVRKSIK-KKRALRRFL-ETELVEPLTPSGTAPNOA | 708 | | |
| QY | 711 | QMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE | 770 | | |
| Db | 709 | QLRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE | 768 | | |
| QY | 771 | AYVMAGVGSYVSRLLGICLTSTVOLVTQLMPYGCLLDHYRNRGLSGDOLLNWCQOIA | 830 | | |
| Db | 769 | ALIMASNDHPLVRLVLLGVCLSPITQLVTQLMPHGCLLEYVHEKDNIGSOLLNWCQOIA | 828 | | |
| QY | 831 | KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGGKVPKWMMA | 890 | | |
| Db | 829 | KGMNLEERLVRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGGKVPKWMMA | 888 | | |
| QY | 891 | LESILRRFRTHQSDVMSYGVVIMELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICTID | 950 | | |
| Db | 889 | LECILHRKFTHQSDVMSYGVVIMELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICTID | 948 | | |
| QY | 951 | VYIMVKNMIDSECRPRFRELSEFSRMDARDPQRFVVIQNEO-LGPASPLDSTFYRSLL | 1009 | | |
| Db | 949 | VYIMVKNMIDSECRPRFRELSEFSRMDARDPQRFVVIQNEO-LGPASPLDSTFYRSLL | 1008 | | |
| QY | 1010 | ZDDMGDLVDAEYLVPOQGFPCPDPAAGAGGVHHRHRSSTRSGGGDLTLGLEPSEEE | 1069 | | |

| | | | |
|--|------|--|------|
| Db | 1009 | DEEDLEMDMAEYLVQAFNIPPP-----IYTSRARIDSNRS-----EIGHSEPPPAY | 1056 |
| Qy | 1070 | APRS-----PLAP-SEGAGSDVFDGDLGMAAKGLQS | 1100 |
| Db | 1057 | TPMSGNQFYVRDGGFAAEQGVSVVPYRAPTSTIPEAPVAQGATAETFDSCCNGTLRKPA | 1116 |
| Qy | 1101 | LPTHDSPLQRYSEDTVPPLPS-----ETQGYAPLTCSPQPEYVNOQDVRPQPPSPR | 1153 |
| Db | 1117 | PHVQEDSSSTQRYSAADTVFAPERSPRGELDEEGYMTFMRDKPKQEYLNPE----- | 1167 |
| Qy | 1154 | EGPLPAARFAGATLERAKTLLSPQKGVKVDVFAFGAVENPEYLTPOGGAAPQPHPPPA- | 1212 |
| Db | 1168 | ENPFVSR-----KNGDLQ-----ALDNPEYHNASNG-----PPKAE | 1199 |
| Qy | 1213 | -----FSPAFDNLYYWDQDPPPERGA--PPSTF | 1237 |
| Db | 1200 | DEYVNEPLYLNTFANTLGAEYKKNILSMPEKAKAFDNDPYNHSLPRSTLQHPDYL | 1259 |
| Qy | 1238 | KGTP-----AENPEYL | 1249 |
| Db | 1260 | QYSTKYFYKQNGRIRPIVAENPEYL | 1285 |
| RESULT 8 | | | |
| S06142 | | | |
| protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish | | | |
| N;Alternate names: epidermal growth factor receptor homolog; kinase-related trans | | | |
| C;Species: Xiphophorus maculatus (southern platyfish) | | | |
| C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000 | | | |
| C;Accession: S06142; S13809 | | | |
| R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Robe | | | |
| Nature 341, 415-421, 1989 | | | |
| A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing | | | |
| A;Reference number: S06142; MUID:90015140; PMID:2797166 | | | |
| A;Accession: S06142 | | | |
| A;Molecule type: DNA | | | |
| A;Residues: 1-1166 <WIT> | | | |
| A;Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291 | | | |
| R;Adam, D.; Maeueller, W.; Scharlt, M. | | | |
| Oncogene 6, 73-80, 1991 | | | |
| A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in X | | | |
| A;Reference number: S13807; MUID:91125882; PMID:1846957 | | | |
| A;Accession: S13809 | | | |
| A;Status: preliminary; translation not shown | | | |
| A;Molecule type: DNA | | | |
| A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA> | | | |
| A;Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285 | | | |
| C;Genetics: | | | |
| A;Gene: mrk | | | |
| A;Map position: Y | | | |
| A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1 | | | |
| C;Superfamily: epidermal growth factor receptor; protein kinase homology | | | |
| C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane prote | | | |
| F;1-25/Domain: signal sequence #status predicted <SIG> | | | |
| F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <M | | | |
| F;707-972/Domain: protein kinase homology <KIN> | | | |
| F;715-723/Region: protein kinase ATP-binding motif | | | |
| Query Match 39.4%; Score 2684; DB 1; Length 1166; | | | |
| Best Local Similarity 45.2%; Pred. No. 1.4e-102; | | | |
| Matches 573; Conservative 165; Mismatches 391; Indels 138; Gaps 28; | | | |
| Qy | 4 | AALCRGILLALLPPCAAST-----QVCTGTDKMLRPLASPETHLDMLRHLVQGCQVVOGN | 59 |
| Db | 8 | ALLQ--LLLVLSISRCCSTDPDRKVCQGTSTNQMTM---LDNHYLKMKNYSGCNVYLEN | 62 |
| Qy | 60 | LELTPLTNASLSFLDIOEVQGVYLIAHNOVROVPLQRLRIVRGTOQLFEDNYALAVLDN | 119 |
| Db | 63 | LEITYTOENQDLSFLOSIEVGVYLIANNEVSTIFLVNLRILRGQNLVEGNTFLVMSN | 122 |
| Qy | 120 | GDPNNTTPTVTGASPGGLRELQRLSLEILKGGVLIRNPQLCYQDTILWKDQYIKANSK | 179 |

Db 123 YQK-NPSSP--DVYQVGLKQLQLSNLTETLSGKGVKSHNPLLCNVETINWMDIVDKTSNP 179
QY 180 FIGITELNRSRACHPCSPCKGSRGWGSEDCQSLTRTVYACGGC-ARCKGPLPTDCCHC 238
Db 180 TWMNLPHAFEROCOKCDHGCWNGSCWAPGPHGCHQKFTKLLCAEQCNRRCRGPKPIDCCNE 239
QY 239 QCAAGCTGPKHSDCLAHFNHSGICELHCPALVYNTDTFESMPNPBGRYTFGASCVTA 298
Db 240 HCAGGCTGPRATDCLACRDNDDGTCKTCTPPKIYDIVSHQVVDVNPNIKTFTGAACVKE 299
QY 299 CPVNYLSTDVGSCTLVCPHLNQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVAVTS 358
Db 300 CRSNVVUTE-GACVRSASAGLEV-DENGKRSCKPGCVKVCDDIGIGSLNNTIAVNS 357
QY 359 ANIQEFAGCKIFGSLAFIPESPDGPASNTAPLQPEQLQVFTLEETITGYLISAWPDS 418
Db 358 TNIRSFNSCTKINGDIILNRNSPEGDPHYKIGTMDPEHLWNLTTVKEITGYLVMWPN 417
QY 419 LPDLSVFONLVRIGRLHNGAYS-LTLOGLGISWGLSLRSLRGSLALIHNTLFCV 477
Db 418 MTSLSVFQNLRIIRGTTTSRSGFSFVVQVVRHQLWGLSLRSLKSEVSNVILKNTLQRYA 477
QY 478 HTVPMDQLFNPHQALLHTANRPEDECEVGEGLACHOLCARGHCWGPGTQCVNCSQFLRG 537
Db 478 NTINRRLLFRSEDQSIYDART-----ENQTCNNECEDGCM-PGPTWCVSLHVDRG 529
QY 538 QBCVEECRVLQGLPRYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFCV 597
Db 530 GRCVASCNLLQGEPREAQVDRGVCQCHQCLVQDLSLTCYGPANCKSAHFQDPQCI 589
QY 598 ARCPGSKPDLSPMPTWKPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTSIVA 657
Db 590 PRCPHGILGDDTL-LWKYADKMGQCPQCHNQTCGSGPGLSGRCD-IVSHSLAVGL 647
QY 658 VVGILLVVLGVVFGVILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAOMRLKE 717
Db 648 VSGLLITVALLIVLLRRRIK-RKTRICLLQELVEPLTPSGQAPNQAFRLKE 706
QY 718 TELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAKVLRNTSPKANKELDEAYVMVAG 777
Db 707 TEFKRDVLGSAFGTVYKGLWNPDCENIRIPVAKVLRNTSPKYNQEVLEAYVMASV 766
QY 778 GSPYVSRLLIGLSTVQLVTLQMPYGCCLLDHVRNRLGSGODLLNWCQIAKMSYLE 837
Db 767 DHPHVCRLGLICLTSVAQLVTLQMPYGCCLLDVYRQHERICGQWLNNWCQIAKGNYLE 826
QY 838 DVRLVHRDLAARNVLKSPNHVKITDFGLARLLIDETEHADGGKVPKMWALESIILR 897
Db 827 ERHLVHRDLAARNVLKSPNHVKITDFGLSKLLTADKKEYQADGGKVPKMWALESIILQ 886
QY 898 RETHOSDVMYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMVK 957
Db 887 TYTHQSDVMYGVTVWELMTFGSKPYDGIPIAKEIASVLENGERLPQPPICTIEVYMIILK 946
QY 958 CMWIDSECPRELVESESRMARQORFVITQNEDLGPASPLDSTFFYSRLLEDMDGDL 1017
Db 947 CMWIDPSSPRFRELVEGFSQMDPSRYLVLTQ-----NLPISLSDRLFRSLSSDD--DV 1001
QY 1018 VDAEYLVLPQOQFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEARPSPLAP 1077
Db 1002 VDAEYLLPYKXI-----NRQGS-----EPCIP 1024
QY 1078 SEGAGSDVFDGLGMAAKGLOSPLTHDPSPLQRYSEDTV-PLPSETDGYVAPLTCSPQ 1136
Db 1025 PTGH-----PVRENSITLRNISDPTQNALEKLDGH-----1055
QY 1137 PEYVNPQVRPOP-----PSPRE-----GLP-AARPAGATLERAKTLSPGKNGVKD 1183
Db 1056 -EYVNPQGETSSRLSDIYNPYEDLTDCGWGPVSLSSQEAETNFRPEYVLTNQNLSL---1111
QY 1184 VFAFGAVENPEYLTPOGGAAPQPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPA 1243
Db 1112 PLVSSGSDDDPY---QAG-----YQAAF-----LPQTGALTGCMFLPAA 1149

QY 1244 ENPEYLG 1250
Db 1150 ENLEYLG 1156

RESULT 9

A36223

kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human

C;Species: Homo sapiens (man)

C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C;Accession: A36223; 159164

R;Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A;Title: Isolation and characterization of ERBB3, a third member of the ERBB/epiderma

A;Reference number: A36223; MUID:90083234; PMID:2687875

A;Accession: A36223

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1342 <KRA>

A;Cross-references: GB:M29366

R;Plozman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro,

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A;Title: Molecular cloning and expression of another epidermal growth factor receptor

A;Reference number: I59164; MUID:90311312; PMID:2164210

A;Accession: I59164

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A;Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841

C;Genetics:

A;Gene: GDB:ERBB3; HER3

A;Cross-references: GDB:119880; OMIM:190151

A;Map position: 12q13-12q13

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hor

C;Keywords: ATP; phosphotransferase

F;707-972/Domain: protein kinase homology <KIN>

F;715-723/Region: protein kinase ATP-binding motif

Query Match 35.6%; Score 2423.5; DB 2; Length 1342;

Best Local Similarity 40.7%; Pred. No. 6.9e-92;

Matches 533; Conservative 191; Mismatches 458; Indels 129; Gaps 32;

QY 10 GLLLALLPGAA--STQVCTGTDMKLRPLASPETHLMDLRHLHYQGVVQGNLELTPLT 67
Db 11 GLLFSLARGSEVNGSQAQVCPGLTGLSLVGTDAENQVOTLYKLYERCEVWGNLEIVLTGH 70
QY 68 NASLSPLODIQEVQGVYLIAHNOVROVPLQRIVRGTQLFEDNYALAVLNDGDLNMTT 127
Db 71 NADLSFLQIREVTGYVLVAMNEFTLPLNLRVVRGTQVYDGKFAIFVM-----LNYNT 125
QY 128 PVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDQYIKANSKFIGITELN 187
Db 126 ----NSSHALRQLRLTQLTEILSGGVYIEKDKLCHMDTIDWRDRIVRDRDAE---IVVKD 178
QY 188 RSRACHPCSPMCKSRGCSSEDCQSLTRTVYACGGC-ARCKGPLPTDCCHCCAAAGCTG 246
Db 179 NGRSCPPCHEVCKG-RWGPGEDEQTLTKTICAPQCNHCHGCPNPQCHDCGACGCG 237
QY 247 PKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPBGRYTFGASCVTAAPYNYLST 306
Db 238 PQDTCFACRHNDSGACVPRCPQPLVYVNTLTFQLEPNPHTKYQYGVGVASCPHNFV-V 296
QY 307 DVGSCTLVCPHLNQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVAVTSANIOEFAG 366
Db 297 DQTSVCVRACPPDKMEVD-KNGLKMCPECGGLCPKACEGTGSG--SRFTVDSSNIDGFVN 353
QY 367 CKKIFGSLAFIPESPDGPASNTAPLQPEQLQVFTLEETITGYLISAWPDSLPDLSVFO 426
Db 354 CTKILGNLDFLTGLNGDPWHKIPALDPEKLVNFRVREITGYLNTQSPPPHMHNFVSFS 413
QY 427 NLQVIRGRILHNGAYS-LTLOGLGISWGLSLRSLRGSLALIHNTLHLCFVHTVPWDOL 485


```
Db 414 NLTTIGRSLYNRGFSLLIMKNLNVTSLSGFRSLKEISAGRIYISANRQLCVHHSLNWTKV 473
Qy 486 FRNPQALLHTA-NRPEDECVGEGLAHQI-CARGHCWGCPPTQCVNCSQFLRGQECVSEC 544
Db 474 LRGPTEERLDIKHNRPRDCVAEGKVCDPICSSGGCGWPGGQCLSCRNYSRSGVCVTHC 533
Qy 545 RVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCAVCAHYKDPDFCVARCPGV 604
Db 534 NFLNGEPREFAHEAEFSCHEPCOMEGTATCNGSGSDTCAQCAHFRDGHPCVSSCPGV 593
Qy 605 KPDLSPYMPIWKFDEBAGAQPCPINCTHSCVDLDDKGCAPAEORA-----SPLTSIVSAVG 660
Db 594 LG--AKGPIYKYPDVQNECRPCHENCCTGGCKGPELQCLGQTLVLIGKTHLTALTVIAG 651
Qy 661 ILLVVLGVVFGILIKRRQKIR-KYTWRRLLQETELVEPLTPSGAMPNOAQWRILKETE 719
Db 652 --LVVIFMILGGTFLYWRGRRIQNKRAMRYLGERGESIEPLDPS-EKANKVLARIFKETE 708
Qy 720 LRKVKVLGSGAFGVYVKGWIPDGENVKI PVAIKVLRNTPSPKANKELIDELVYVMAVGVS 779
Db 709 LRKLVLGSGVGTVHKGVWIPGESIKIPVICIKVIEDKSGRQSFQAVTDHMLAIGSLDH 768
Qy 780 PYVSRLIGICLTSTVQLVTLQMPYVGCILLDVHRENRLGSGDQLLNWCMIKAGNSYLEDV 839
Db 769 AHIVRLGLCPGSSQLVTVQLPLGSLLDHVQRHGAIGPQLLLNMGVQIAKGMYYLEEH 828
Qy 840 RLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRF 899
Db 829 GMVHRNLAARNVLKSPQOVADFVADLLPPDDKQLLYSEAKTPIKMMALESIHFGKY 888
Qy 900 THQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGBRLPQPICTIDVYMIWVKW 959
Db 889 THQSDVMSYGVTVWELMTFCAEPYAGLRLAEVDPDLLEKGERLAQPQICTIDVYMIWVKW 948
Qy 960 MIDSECPRELVSERMRARDQRFVITQNEDLGPA---SPLDSTFYRSLLEDDMDGD 1016
Db 949 MIDENIRPTEKELANEFTWARDPPLYLVTKRES-GPGIAPGPEPHGLTNKKLEVEELEP 1007
Qy 1017 LVDAEYLVPOQGFCDPAPAGAGWVHRHRSSTRSGGDLTLGLPEP-SEEAAPSPL 1075
Db 1008 ELDDLDEAED-----NLATTLGSAISLPGVTLNRRPGSQSL 1048
Qy 1076 APSEGAGSDVFDGLGMAAGLQSLPHD-PSPLQRYSEDPTVPLP-----SETDGYV 1128
Db 1049 SPSSGY-MPMNQNLGSCQESAVSGSERCPRPVSLH-----PMRPGCLASSESSGHV 1101
Qy 1129 A-----PLTCSPOPE-----YVNPQDVRPQPPSPREP-----L 1157
Db 1102 TGSEAELOEKVSMCRSRSRSPRGRGDSAYHSQRHSLLTPTVPLSPGLEEEDVNGYV 1161
Qy 1158 PAARPAGATLERAKTLP-GKNGV-----KDVFAFGAVENPEYLTPOGGAPOPHPP 1210
Db 1162 PDTHLKTPSRSREGLTSSVGLSSVLGTEBED-----EVEYWNRRRRHSPP-PHPP 1212
Qy 1211 PAFSPAFDNIYYWD-----QDPPERGAPPTFKGTPTAENPEYL 1249
Db 1213 RPSSEELGVEYMDVGDLSLGSSTQCPLHPVPIPTAGTTDEDEYEM 1263
```

RESULT 10

```
JC4387
epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A;Reference number: JC4387; MUID:96096535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>
```

```
A;Cross-references: GB:U29339; NID:g915389; PID:g915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for resic
C;Comment: This protein is a functional heregulin receptor that transduces signals to
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hon
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F;640-659/Domain: transmembrane #status predicted <TM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1212,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)
```

```
Query Match 34.4%; Score 2339.5; DB 2; Length 1339;
Best Local Similarity 40.8%; Pred. No. 1.9e-88;
Matches 523; Conservative 172; Mismatches 433; Indels 155; Gaps 35;
```

```
Qy 3 LAALCRWGLLALLPPCAA---STQVCTGTDMLRLPASPTHLDMLRLHLYQGCVVQGN 59
Db 7 LQVLC-----PLLSLRGSEMGNSQAVCPGTLNGLSVTGDADNQVQTLYKLYKECEVWNG 62
Qy 60 LELTYLPTNASLSFLQDIQEVQGVYLIANHVRQVLRIRVRGTQLFEDNYALAVLDN 119
Db 63 LEIVLTCHNADLSFLQWIREVTAYLVAMNEFSLPLNLRVVRGTQVYDGKFAIFVM-- 120
Qy 120 GDPINNTPTVGASPGRLRELRLSRLTEILKGGVLIOJRNPOLCVQDITLWKQVYKANSK 179
Db 121 ---LNYNT----NSSHALRQLKFTLTELISGGVYIEKDKLCHMDITDWRD-IVRVGA 172
Qy 180 FIGITELNRSRACHPCSPCKGSRCSSESSEDQSLTRTVCAAGC-ARCKGPLTDCCHE 238
Db 173 EIVVK--NNGANCPCHVECKG-RCWGGPGDQQLITKTCAPOCNCRGCFGNPQCCHD 229
Qy 239 QCAAGCTGPKHSCLACLHFNHSGICELHCPALVYNTDTFESNPNEGRTYFGASCVTA 298
Db 230 ECAGGCGPQDTCFACRRENDSGACVPRCPPELVYVKNLTFQLEPNPHTKYQGVGVAS 289
Qy 299 CPNYLSTDVGSCTLVPLHNQVETADGTQRECKSKPCARVCYGLGMEHLREVRVTS 358
Db 290 CPNHFV-VDQTFVCRACPPDKMEVD-KHGLKMECPGGLCPKACEGTGSG--SRVQTVDS 345
Qy 359 ANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPQLOVFFETLEITGYLYISANPDS 418
Db 346 SNTDGFVNCTKILGNLDFLITGLNVDPHKIPALDPEKLVNFRVTRITGLNLTQSNPPH 405
Qy 419 LPDLSVFQNLQVIRILHNGAYS-LTLOGLGISWLSRLSRLRELGSGLAIHHNTHLCFV 477
Db 406 MHNFSVFNLTITIGRSLYNRGFSLLIMKNLNVTSLSGFRSLKEISAGRVYISANQQLCYH 465
Qy 478 HTVPWDLFRNPQALLHTA-NRPEDECVGEGLAHQI-CARGHCWGCPPTQCVNCSQFLR 536
Db 466 HSLNWTLLRLLGPSEERLDIKYDRPLGELAEKGVCDPLCSSGGCGWPGGQCLSCRNYSR 525
Qy 537 GQSCVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCAVCAHYKDPDFC 596
Db 526 EGVCTVHCNPLQGEPRFVHEAQCFSCHPECLPMWGTSTVNGSGSDACARCAHFRDGHPC 585
Qy 597 VARCPSGVKPDLSPYMPIWKFPPDEEGACQPCPINCTHSC--VDLDDKGCAPAEQASPLTSI 654
Db 586 VNSCPHGILG--AKGPIYKYPDQAQNECRPCHENCCTQCGNGPELQDCLGQAEVLMSKPHLV 643
Qy 655 VSAVGIILLVVLGVVFGILIKRRQKIR-KYTWRRLLQETELVEPLTPSGAMPNOAQMR 713
Db 644 IAVTVG--LAVILMILGGSFLYWRGRRIQNKRAMRYLGERGESIEPLDPS-EKANKVLAR 700
Qy 714 ILKETELRKVKVLGSGAFGVYVKGWIPDGENVKI PVAIKVLRNTPSPKANKELIDELVYV 773
Db 701 IFKETELRKVLGSGVFGVIVHIGWIPEGESIKIPVICIKVIEDKSGRQSFQAVTDHMLA 760
Qy 774 MAGVSPYVSRLIGICLTSTVQLVTLQMPYVGCILLDVHRENRLGSGDQLLNWCMIKAGM 833
```



```
QY 697 VEPLTPSGAMPNOARILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVL 756
Db 114 VEPLTPSGEAPNQHRLILKETEFKVKVLGFGAGFTVYKGLWPEGEKVTIPVAIKELR 173
QY 757 ENTSPKANKEILDYAYVMAGVSPYVSRLLGICLTSTVQLVTQLMPLYGCLLDHVRENKR 816
Db 174 EATSPKANKEILDYAYVMASVDNPHVCRLLGICLTSTVQLITQLMPLYGCLLDYIREHKN 233
QY 817 LGSQDLLNMCQIAKGNLYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQDLLNMCVQAKGNLYLEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKE 293
QY 877 YHADGGKVPKWMALLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLE 936
Db 294 YHAEGGKVPKWMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353
QY 937 KGERLPQPPICTIDVYIMVKWMIDSECRPRELYSEFMRARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPPICTIDVYIMVKWMSDADSRPKFRELIAEFKWARDPPRYLVIOGDERMH 413
QY 996 PASPLDSTFVRSLLDDMDGLVDAEYLYPQOGFFCPDPAPGAGGMVHRRHSSSTRSG 1055
Db 414 LPSPTDSKFYRLMEEDMEDIVDAEYLYPHQGFF-----NSPST--- 454
QY 1056 GBDLTGLSEPSEEA PRSP-----APSEGAGSDVFDGDLGMGAAGLQSLPTHDPSPLO 1110
Db 455 -----SRTPLLSLSATSNSATNCIDRNG-----H----- 481
QY 1111 RYSEDPTVPLPSETDGYVAPLTCSPQPEYVYQPDVVRPQPPSPREGPLPAARPAAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVYQ--LMPKPSAMVQVQIYNYISLTAISK 523
QY 1170 AKTLSPGKNGWKVDVFAFGGAVENPEYL 1197
Db 524 LPMDSRYQN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:08:35
Job time : 30.9062 secs